

EDITED BY

Maciej Pękala
and
Marek Szepski

Information management in the era of the Internet

SELECTED ISSUES

KRAKÓW 2020

Information management in the era of the Internet

SELECTED ISSUES

EDITED BY

Maciej Pękała
and

Marek Szepski

Information management in the era of the Internet

SELECTED ISSUES



SOCIETAS
VISTULANA

KRAKÓW 2020

Publication is a part of project “KA 2.0 – the development programme of Andrzej Frycz Modrzewski Krakow University” realized as a part of Operational Programme Knowledge Education Development, Axis III Higher Education for economy and development; Measure 3.5. Comprehensive university programmes. The project is co-financed by the European Union within the European Social Fund.



Fundusze Europejskie
Wiedza Edukacja Rozwój



**Rzeczpospolita
Polska**

Unia Europejska
Europejski Fundusz Społeczny



Review:

dr hab. Paweł Wołoszyn

Copyright by Krakowska Akademia im. Andrzeja Frycza Modrzewskiego & Towarzystwo Naukowe Societas Vistulana, Kraków 2020

Edition:

Towarzystwo Naukowe “Societas Vistulana,” www.vistulana.pl

Cover design & DTP:

Justyna Kramarz – Studio Grafpa, www.grafpa.pl

ISBN 978-83-65548-82-5

Printed in Poland



Table of contents

Introduction	7
Maciej Pękala, Marek Szepski	
Information-based civilization – an attempt at definition in the light of management sciences	9
Maciej Pękala	
Linguistic accessibility of websites	23
Krystian Kurnik, Wojciech Baran	
Endangered information privacy	37
Renata Uryga	
The use of “Check Alert” technique in data quality management and business analysis within the company ...	49
Maciej Pękala, Marian Krupa	
Individual information management. Personal digital archives	71
Aneta Januszko-Szakiel	
Information security management – audit of the IT system	83
Mariusz Grzyb, Dorota Kowalik	
Prioritizing areas of implementation within the dynamic modelling concept as regards implementation processes of integrated systems	105
Piotr Komsta	



Introduction

We are pleased to present a publication that was created following the first conference in the “IT and society” series, which took place on 15 November 2019 at the Faculty of Management and Social Communication at the Andrzej Frycz Modrzewski Krakow University.

The publication contains the transcripts of the conference participants’ lectures in particular. It is a review of the role played by information in managing various areas of human activity, and of the related academic interests of the conference participants. The authors also indicate practical applications of the results of the above theoretical considerations.

The following issues can be clearly noted in this monograph: information-based civilization, i.e. defining civilization through information, the discussion of the concept of information, language as an information carrier, the role of information in private life and in business, information security, archiving and management. As the issues discussed are very extensive, the individual chapters are more of an introduction than the final solution to the problem. This specifically applies to the chapters referring to theoretical considerations.

It is a truism that the information transformed into knowledge is a key element of management that allows for making most reasonable decisions. The volume starts with the considerations on the role performed by information in the society and a search for an answer to the question whether our civilization is an information civilization and what it means.

The authors of the next chapter deal with the issue of automatic website translation and its effectiveness. The authors makes an attempt at finding an answer to the question why certain websites are translated automatically and others are not.

The next chapter deals with the issue of our privacy with the vast amount of information currently available on the Internet. The author is looking for answers

to questions about the possibility of controlling what happens with our personal information and whether it is a problem for young people today.

The next chapter offers the visualized results of data analysis, and an original technique of detecting certain types of data errors used for the analysis. Since the accuracy of management decisions based on these analyses is important from an economic point of view, the ability to detect possible data errors will help to avoid making wrong decisions.

Personal digital archives are an important aspect of information management. Each of us stores digital photos, videos and music, one's own texts and texts received from others, and downloaded from the Internet. Maintaining order that allows you to quickly find the information you need is not easy and the author analysed the problems related thereto.

Any information is valuable, therefore you need to keep it secure. The Authors of the next chapter deal with issues related to information security management.

The final chapter focuses on the priorities in the implementation of management support systems. The primary task of these systems is information processing, and the priorities are determined by the importance of the processing effects for enterprise management.

In all considerations, the Authors pay attention to the important role of information in making management decisions. The issues related to the core, processing, security and archiving of information are considered from the point of view of civilization, organization (enterprise) and from a personal point of view.

Editors

Maciej Pękala

Marek Szepski

Maciej Pękała
ORCID No. 0000-0002-4785-3583
Andrzej Frycz Modrzewski Krakow University

Information-based civilization – an attempt at definition in the light of management sciences

Abstract

Two issues are considered in the publication: What is an information civilization? and Are we part of it in Poland now? To answer the question, the analytical-synthetic method was used in the field of scientific and industry literature. As a result of the considerations, it was concluded that the term “information civilization” is well defined. It was also stated that in Poland we are currently on the way to the information civilization. The final conclusion is that the term “information civilization” defines the direction in which today’s post-industrial civilization is heading.

Introduction

Recently, we commonly use the term “information society” previously referred to as “information society”. In this work, I would like to analyse the term “information civilization”. The considerations will be focused on the correctness of terminology as well as on determining whether we are currently part of the information civilization in Poland. We will attempt to answer the following questions:

1. What is an information civilization?
2. Are we part of it currently in Poland?

The analytical and the synthetic method will be used for scientific and industry literature as a research method.

1. Historical connotations

Before dealing with the concepts of “civilization” and “information”, here are a few references to literature. In the 1970s, the Japanese Yoneji Masuda formulated the requirements for the existence of an information society in Japan. The results of the research on this topic were published in the 1980 book titled “Information Society as Post-Industrialized Society”.

The following Table 1 was placed there:

Table 1. The developmental stages of computerization.

Stage of development	First stage 1945–1970	Second stage 1955–1980	Third stage 1970–1990	Fourth stage 1975–2000
Bases of computer usage	big science	management	society	individual
Goal	national defence, space exploration	gross national product GNP	gross national welfare GNW	gross national satisfaction (GNS)
Values	national prestige	economic growth	social welfare	self-actualization
Subject	nation	organizations	general	individual
Object of computer use	nature	organization	society	human beings
Scientific base	natural sciences	management sciences	social sciences	behavioural sciences

Stage of development	First stage 1945–1970	Second stage 1955–1980	Third stage 1970–1990	Fourth stage 1975–2000
Information object	attaining scientific goals	pursuing business efficiency	solving social problems	intellectual creation

Source: Y. Masuda, *Information Society as Post-Industrialized Society*, Tokyo 1980, p. 37.

This table relates to the project of creating information society in Japan¹. In the above publication by Masuda, we will also find the following text: “(...) the civilization to be built as we approach the 21st century will not be a material civilization symbolized by huge constructions, but will be virtually all invisible civilization. Precisely, it should be called «information civilization». Homo sapiens, who stood at the dawn of the first material civilization at the end of the last glacial period, is now standing at the threshold of the second, the information civilization after ten thousand years”². Here, for the first time, we find the concept of information civilization. Its features are defined in the last column of Table 1. The term can also be found in Kisielnicki’s book: “It is not the problems of consumption but the problems of information and knowledge that are at the focus of our attention. Modern civilization is often referred to as information civilization. In this new society, information comes to the forefront. A person becomes free because they have information and knowledge, and these resources allow the person to decide about their own fate”³.

The term “information society” is earlier, it is dated to 1963, and the author is Japanese – Tadao Umesao. The data in this respect can be found in the article by Moryś⁴. Since 1963, many definitions of the term “information society” have appeared. The outline can be found in the article: by Nowak⁵, and their discussion in the text by Buregwa-Czuma and Garwol⁶.

¹ Y. Masuda, *Information Society as Post-Industrialized Society* [1980], p. 156. Cf. T. Goban-Klas, *Cywilizacja medialna. Geneza, ewolucja, eksplozja*, WSIP, Warszawa 2005, p. 38.

² Y. Masuda, *Information Society as Post-Industrialized Society* [1980], p. 156. Cf. T. Goban-Klas, *Cywilizacja medialna*, p. 38.

³ J. Kisielnicki, *Systemy informatyczne zarządzania* (ebook), Wydawnictwo PLACET, 2014, p. 413.

⁴ A. Moryś, *Geneza i ewolucja wywiadu gospodarczego. Część pierwsza*, “Infotezy”, 1 (2011), 1; <http://www.ujk.edu.pl/infotezy/ojs/index.php/infotezy/issue/view/2> (accessed on August 15, 2019).

⁵ J.S. Nowak, *Spółczesność informacyjna – geneza i definicje*, 2nd edition, revised and extended 2008; http://delibra.bg.polsl.pl/Content/24702/BCPS_28204_2008_Spoleczenstwo-inform.pdf (accessed on August 15, 2019).

⁶ S. Buregwa-Czuma, K. Garwol, *Definicje, właściwości i funkcje społeczeństwa informacyjnego*, “Dydaktyka Informatyki”, 6 (2011), pp. 30–37.

The question of the term “information” or “IT” still remains to be clarified. In the paper by Kuraś⁷ we can find the following definitions:

“**Information system** is a social system («human activity system»), which is made up of elements belonging to five classes: **data**, **methods**, **technique** (technology used – technical equipment), **organization**, **people**”⁸ and “**IT system** is a set of techn [ical] and log [ical?/istic?]] used to collect, store and process information (PWN [2004])”⁹.

The paper thoroughly discusses the differences between these concepts. The most important difference is that the term “informational” is related to the processing of information by a human being, and to the term “computerized” to the processing of data by a computer. Thus, the correct term is information society.

2. Civilization – definition and characteristics

In the online PWN Dictionary of the Polish Language¹⁰ we can find the following definitions of the term “civilization”:

1. “The society’s state of development in a given historical period, determined by the degree of mastery of nature by man”.
2. “All material goods, means and skills achieved by a given society in a given historical era”.

In the online PWN Encyclopaedia, on the other hand¹¹, we read:

“Civilization [Latin: *civilis*], the level of development achieved by society in a given historical epoch, with particular emphasis on the level of material culture (especially science and technology), which is an indicator of human mastery of the forces of nature and the use of its resources”.

In the subsequent part of the slogan, it is pointed out that the term “civilization” is used in three meanings: historical; type of culture advanced socially, materially and ideologically; a set of ethnic cultures that together form a cultural circle. In our

⁷ M. Kuraś, *System informacyjny a system informatyczny – co oprócz nazwy różni te dwa obiekty?*, “Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie / Cracow Review of Economics and Management”, 770 (2009), pp. 259–275.

⁸ Ibid., p. 265.

⁹ Ibid.

¹⁰ <https://sjp.pwn.pl/slowniki/cywilizacja.html> (accessed on August 18, 2019).

¹¹ <https://encyklopedia.pwn.pl/haslo/cywilizacja;3888917.html> (accessed on August 18, 2019). The same definition is also available in the printed version: *Nowa Encyklopedia PWN*, vol. 1: A-C, PWN, Warszawa 1998.

considerations, we are interested in the second meaning, consistent with the content of the definition. It is primarily about the technological and the cultural aspect. It is found in civilizations from the nineteenth century to the twenty-first century, e.g. scientific and technical, industrial or post-industrial civilization. It is worth noting that in a given historical period, there are several parallel definitions of civilization, if only because of the three meanings mentioned. Following Masuda's ideas, we still have to answer the question whether the post-industrial civilization is an information civilization? Before attempting to answer this question, we will look at the adjective "informative".

3. Information – definition and characteristics

Let us start with the most general definition in the online PWN Encyclopaedia¹²:

Information [Latin: *informatio* "image", "explanation", "notification"]:

1. concept, in principle undefinable due to its primary, elementary nature; most frequently analysed in three aspects: syntactic (concerns the amount of information that can potentially be included in a given message), semantic (meaning and content of the message) and pragmatic (usefulness of the information, i.e. the value of the information contained in the message due to the goal pursued by the recipient). There are currently no universal methods of information analysis known in the semantic and pragmatic meaning of the term. From the syntactic point of view, information is defined either by the amount (measure) of information I (information theory), or as a synonym for the term of datum (data). In the theory of information $I = \log(1/p) = -\log p$ where p is the probability of the occurrence of an event (or a sequence of events) from a set of possible events, where the event is for instance the presence of a certain element of the set – the occurrence of a certain character or letter at a given position of the message being transmitted; the unit of quantity of information is shannon.
2. in everyday language, an observation of the state of affairs, a message. The term information is also related to the media, in particular to the press¹³ and is the subject of information theory research¹⁴. In the PWN Encyclopaedia we can find the following definition of information theory: "the theory of transmitting messages from the source of the news to the object of their destination – the

¹² <https://encyklopedia.pwn.pl/haslo/informacja;3914686.html> (accessed on August 19, 2019). Cf.: *Nowa Encyklopedia PWN*, vol. 3: I-L, PWN, Warszawa 1997.

¹³ <https://encyklopedia.pwn.pl/haslo/informacja;4008526.html> (accessed on August 19, 2019).

¹⁴ <https://encyklopedia.pwn.pl/haslo/informacja;3914687.html> (accessed on August 19, 2019).

recipient (mouth); in a broader sense, also a science dealing with a comprehensive analysis of the processes of transmitting information using sign systems in society and in biological communities”.

Claude Shannon’s information theory became the basis for information transmission and storage techniques¹⁵, not only in the sphere of theory, but also in the construction of commonly available devices. The second element is the development of electronic, semiconductor and optical technologies.

4. A combination of civilization + information

In the definition of the term “civilization”, one of the elements is the level of exact knowledge and technology, and this is an indicator of human mastery of the forces of nature and the use of its resources. Currently, there is no area of human activity where a computer, being the basic tool for processing data into information, would not be used. Today we constantly face information technologies. Among other things, Big Data processing techniques and the Internet of Things (IoT) is slowly entering our homes and workplaces. We may thus consider information technologies as the dominant ones. The extensive use of computer science in science, technology and everyday life entitles us, according to the definition of civilization, to create the term “information civilization”.

5. The characteristics of information civilization

Let us consider the characteristics – the determinants of information civilization. I will start by quoting the table with the Masuda Plan (Table 2).

¹⁵ Cf. J. Gleick, *Chaos. Narodziny nowej nauki*, Wydawnictwo Zysk i S-ka, Poznań 2018, p. 318.

Table 2. Masuda's plan.

	Phase I 1945–1950 Computerization of great science	Phase II 1950–1970 Computerization of management	Phase III Plan 1970–1980 Computerization of social information	Phase IV Plan 1980–2000 Computerization of individual activities
Purpose	defence, development, space research	Gross National Product	prosperity, social welfare	satisfaction
Scale of valuation	national prestige	economic growth	social welfare	personality development
Entity	country	enterprise	population	individual
Subject	nature	organization	society	human entity
Basic science	discipline, natural science	management sciences	social sciences	learning about individual's behaviour
Information pattern	achieving the goal	effectiveness	problem solving	intellectual creativity

Source: T. Goban-Klas, *Cywilizacja medialna*, p. 38.

Let us note that the same determinants may be used to check whether we deal with information civilization in the social field, yet in civilization it is important to use knowledge and technology. We will look at Figure 1 taken from the work of Masuda:

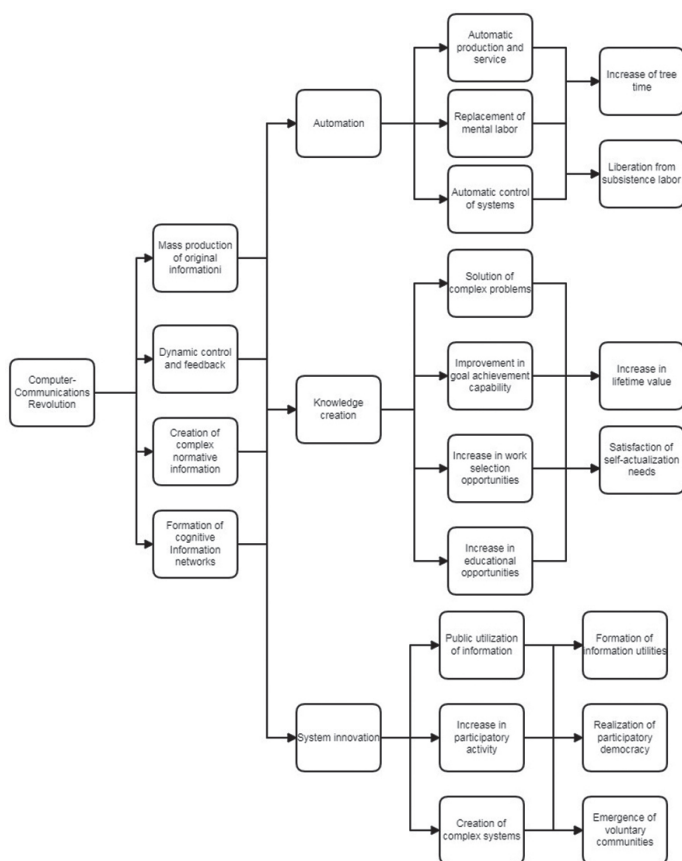


Figure 1. Computer communications revolution and its societal impact. comp.

Source: based on Y. Masuda *Information Society as Post-Industrialized Society*, p. 61.

The graph illustrates the significant influence of the computer on such spheres of human activity as production automation, replacing mental work – this may be what we should worry about, knowledge creation, public use of information, increase in participatory activity. Since 1980, there has been a surge in computing power, mobile computers and 3D printers have appeared, and this application of information technology has entered many areas previously unavailable. Examples include three-dimensional imaging and robots used in medical operations, digital radio and television, autonomous cars and drones, near-autonomous passengers and transport aircraft, artificial intelligence systems to support and even make decisions, analyse data collected from astronomical observations and physical experiments and can still be replaced. Let us also look at Figure 2 illustrating civilization transitions.

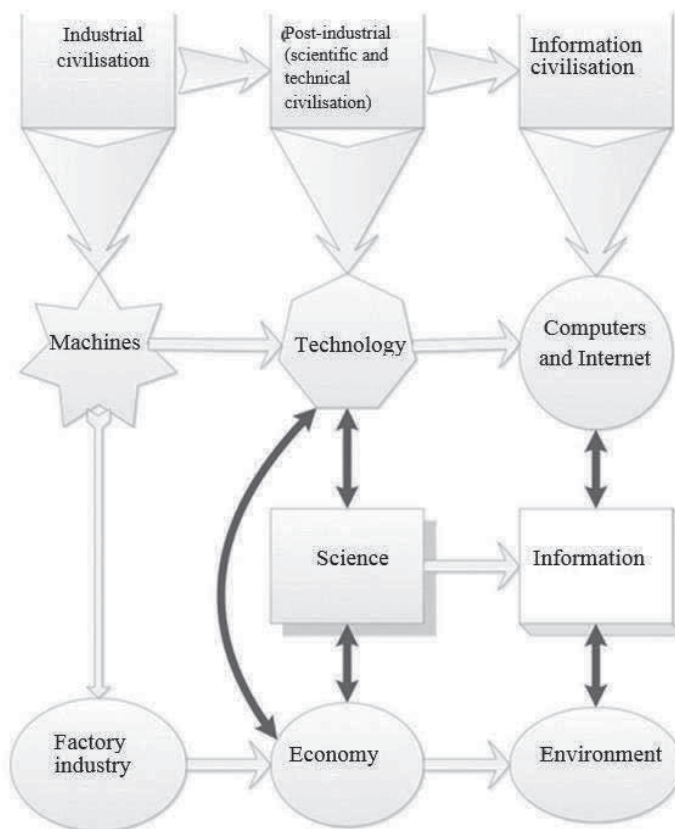


Figure 2. Civilization transitions.

Source: own elaboration.

Apparently, the shift from machinery and industry will support science and technology towards computerization and information. So let us consider the distinguishing characteristics of the information civilization:

- computers and networks;
- automation of activities previously performed only by humans;
- solving complex problems;
- more time to develop your own personality;
- gathering information.

We have features, but no measure, but we can answer the question whether Poland is part of the information civilization. I will sketch the answer, leaving the broader justification for a separate text.

Computers and networks – including access to the internet, are commonplace in every category, from supercomputers to mobile devices. Automation of activities previously performed only by humans – we have a wide range of solutions here, from industrial robots, through medical implants, intelligent everyday devices, to activities in management systems.

Solving complex problems – the use in scientific research, the use of artificial intelligence and simulation models, and in management Business Intelligence systems. More time to develop your own personality – easy access to movies, books, games via mobile devices, but also to current knowledge in any field. Gathering information – databases for the needs of organizations, public and state administration, as well as photos, videos and other data in the social media.

6. Conclusions and recommendations

Note that the term “information civilization” has been well defined. Although I believe that in the 21st century we live in an information civilization, in this article, due to the shortened justification, I will say that the term “information civilization” defines the initial phase of transformation of the post-industrial civilization.

As the applications of information technologies in Poland do not differ from the global level, and we are pioneers in some areas, it is legitimate to conclude that we are currently on the road to information civilization in Poland.

I believe that the term information civilization can and should be used to describe modern civilization.

Instead of conclusion: a bit of futurology

Data is a natural resource, such as water, coal, wood, gas and oil. The result of data mining is information¹⁶. Information is a commodity¹⁷ as a result of data transformation by IT or information systems.

In the short story by Stanisław Lem, *Professor A. Dońda. From the Memories of Ijon Tichy*¹⁸, Professor Dońda says “(...) information has a mass. Every piece of it. Anything. The content doesn’t matter at all. The atoms are the same whether

¹⁶ Cf. P. Beynon-Davies, *Inżynieria systemów informacyjnych*, WNT, Warszawa 2004, pp. 15–16.

¹⁷ Ibid., p. 42.

¹⁸ A short story in the collection by S. Lem, *Maska*, Wydawnictwo Literackie, Kraków 1976, pp. 51–80.

they are in stone or in my head. Information has a load, but its weight is extremely small”¹⁹ and further “Mass, energy and information are the three forms of mass, he explained patiently. – They can change into each other according to the conservation laws. (...) Matter turns into energy, energy and matter are needed to produce information, and information can pass back into it, of course not just anyhow. Above the critical mass, it is like blown away”²⁰. Professor Dońda has accumulated 490 billion bits of information in a memory set on a sensitive scale, i.e. approx. 57.04 TiB²¹ and the weight showed a weight gain of 0.01 g. How correct is the prediction of S. Lem? John Kubiawicz from the University of Berkeley calculated²² that the eBooks on the Kindle are weighted $1 * 10^{-18}$ g, or 1 attogram. You can see that the data is much lighter than it was suggested by S. Lem. In 2011, the executive director of Google Eric Schmitt estimated the physical weight of data circulating on the Internet²³. There are two basic approaches: device weight and electron mass. More interesting is the calculation based on the electron mass. Eric Schmitt, using Einstein’s formula for the equivalence of mass and energy, converted the energy of an electron to mass, and then multiplied the mass of the electron equal to $9,109 * 10^{-28}$ g by the estimated number of electrons and obtained a value of just over 50 g.

In 2018, the Internet was estimated at 10 YiB²⁴ (yottabytes) of data or $10 * 10^{24}$ and that adds up to a few kilos. Data is growing at a rate of gigabytes per second. Since the known Einstein formula allows you to convert mass into energy, perhaps our great-grandsons, after depleting natural resources, will use a given source of energy. Unless, according to the formula by Prof. A. Dońda, the data (information), after exceeding the critical mass, will turn into energy. If not, then we will have some new civilization that can recover the energy stored in data and will use quantum computers whose theory is developed²⁵, and more recently there have been reports of the first commercial quantum computers²⁶. In October 2019,

¹⁹ Ibid., p. 70.

²⁰ Ibid., pp. 71–72.

²¹ Binary prefixes according to the norm IEC-60027–2.

²² <https://www.pcworld.pl/news/Kindle-jest-ciezszy-gdy-zaladkuje-na-nim-e-booki-serio-serio,377125.html> (accessed on September 3, 2019).

²³ <https://m.niezalezna.pl/222513-ile-wazy-internet> (accessed on September 2, 2019).

²⁴ <https://innpoland.pl/144691,ile-miejsca-zajmuje-caly-internet-ile-wazy-wspolczesny-internet> (accessed on September 2, 2019).

²⁵ The presentation of quantum information theory can be found in e.g. M. Le Bellac, *Wstęp do informatyki kwantowej*, PWN, Warszawa 2012.

²⁶ <http://www.national-geographic.pl/nauka/pierwszy-komercyjny-komputer-kwantowy-wchodzi-na-rynek> (accessed on September 14, 2019).

Google announced its own quantum computer²⁷. So far, mainly mathematical problems have been solved. Currently, there are no universal quantum computers, and in order to solve the problems of a certain class, it is necessary to construct an appropriate quantum computer. You can see some similarity to analogue computers. There is still a long way to process information by quantum computers in terms of management, and perhaps ERP systems will never use quantum computers, which can be useful in BI decision support systems. Google engineers believe that “the era of quantum supremacy is approaching faster and faster, which may take the development of our civilization to an entirely new level”²⁸. I would call it the Quantum Civilization.

References

- Beynon-Davies P., *Inżynieria systemów informacyjnych*, WNT, Warszawa 2004
- Buregwa-Czuma S., Garwol K., *Definicje, właściwości i funkcje społeczeństwa informacyjnego*, “Dydaktyka Informatyki”, 6 (2011), pp. 30–37
- Gleick J., *Chaos. Narodziny nowej nauki*, Wydawnictwo Zysk i S-ka, Poznań 2018
- Goban-Klas T., *Cywilizacja medialna. Geneza, ewolucja, eksplozja*, WSiP, Warszawa 2005
- Kisielnicki J., *Systemy informatyczne zarządzania* (ebook), Wydawnictwo Placet, Warszawa 2014
- Krztoń W., *XXI wiek – wiekiem społeczeństwa informacyjnego*, “Modern Management Review”, 22 (2015), pp. 101–112
- Kuraś M., *System informacyjny a system informatyczny – co oprócz nazwy różni te dwa obiekty?*, “Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie / Cracow Review of Economics and Management”, 770 (2009), pp. 259–275
- Le Bellac M., *Wstęp do informatyki kwantowej*, PWN, Warszawa 2012
- Lem S., *Profesor A. Dońda. Ze wspomnień Ijona Tichego*, [in:] *Maska*, Wydawnictwo Literackie, Kraków 1976, pp. 51–80
- Masuda Y., *Information Society as Post-Industrialized Society*, Tokyo 1980; 2nd edition, World Future Society, Washington 1983
- Nowa Encyklopedia PWN*, PWN, Warszawa, vol. 1: A–C, 1998; vol. 3: I–Ł, 1997
- Nowak J.S., *Spółeczeństwo informacyjne – geneza i definicje*, 2nd edition, revised and extender 2008; <http://delibra.bg.polsl.pl/Content/24702/>

²⁷ <https://www.wirtualnemedial.pl/artykul/google-komputer-quantowy-trzy-minuty-obliczen-zamiast-10-tys-lat-zastosowanie-budowa-specjalizacji-wydajnosci-opinie> (accessed on October 25, 2019).

²⁸ Ibid., last paragraph.

BCPS_28204_2008_Spoleczenstwo-inform.pdf (accessed on August 15, 2019);
(1st edition under *Spółeczeństwo informacyjne 2005*, G. Bliźniuk, J.S. Nowak (ed.),
Wydawnictwo PTI, Katowice 2005)

References (online)

- Moryś A., *Geneza i ewolucja wywiadu gospodarczego. Część pierwsza*, “Infotezy”, 1 (2011),
1; <http://www.ujk.edu.pl/infotezy/ojs/index.php/infotezy/issue/view/2> (accessed on
August 15, 2019)
<https://sjp.pwn.pl/slowniki/cywilizacja.html> (accessed on August 18, 2019)
<https://encyklopedia.pwn.pl/haslo/cywilizacja;3888917.html> (accessed on August 18, 2019)
<https://encyklopedia.pwn.pl/haslo/informacja;3914686.html> (accessed on August 18, 2019)
<https://encyklopedia.pwn.pl/haslo/informacja;3914687.html> (accessed on August 19, 2019)
<https://encyklopedia.pwn.pl/haslo/informacja;4008526.html> (accessed on August 19, 2019)
<https://encyklopedia.pwn.pl/haslo/informacji-teoria;3914691.html> (accessed on August 18, 2019)
<https://encyklopedia.pwn.pl/haslo/wiadomosc;3995309.html> (accessed on August 18, 2019)
<http://www.inzynierawiedzy.pl/wiedza/wiedza-informacje-dane> (accessed on August 18, 2019)
<https://www.pcworld.pl/news/Kindle-jest-ciezszy-gdy-zaladujesz-na-nim-e-booki-serio-serio,377125.html> (accessed on September 3, 2019)
<https://www.wirtualnemedi.pl/artikul/google-komputer-kwantowy-trzy-minuty-obliczen-zamiast-10-tys-lat-zastosowanie-budowa-specyfikacja-wydajnosci-opinie> (accessed on
October 25, 2019)
<https://www.google.com/amp/s/tech.wp.pl/ile-wazy-internet-wynik-jest-zadziwiajacy-6034856791458433a%3famp=1> (accessed on September 2, 2019)
<https://www.google.com/amp/s/innpoland.pl/amp/144691,ile-miejsc-zajmuje-caly-internet-ile-wazy-wspolczesny-internet> (accessed on September 2, 2019)

Krystian Kurnik
ORCID No. 0000-0002-6644-0856
Andrzej Frycz Modrzewski Krakow University

Wojciech Baran
ORCID No. 0000-0003-0692-9702
Pedagogical University of Krakow

Linguistic accessibility of websites

Abstract

Automatic website translation is a great help for users. The option to translate from virtually any language to any other language allows anyone to virtually understand any text published and shared on the web. However, the quality of translation depends on specialized algorithms that allow the translation of complex phrases. It is interesting, however, how the option to click and translate pages is suggested for some pages and not suggested for others. Therefore, the study touched upon websites in terms of, inter alia, technology in which they are created in relation to the existence of a translation proposal or the lack of it.

The article will show the results of the analysis carried out on a several dozen websites made in various technologies, meeting the selected keywords.

Introduction

Knowledge is a common good that we all share, whether in written or spoken form. The expansion of the Internet and the global network made it possible to significantly speed up data flow. The information collected on the websites is used to transfer information. An extensive database created by the websites cannot be limited in any way in the era of digitization and data collection.

Nowadays, the most common language is English. However, websites appear in as many languages as those publishing them on the Internet. Language is not a barrier. Content management systems now have the ability to switch between the selected languages of the website. This is an option implemented by the authors of the website. However, this is sometimes not enough. It happens that the content cannot be translated. How to act in such a case? The option of machine translation has been developed over the years. With the machine translation technology, there is no longer any problem with understanding the data contained on websites.

Review of the selected technologies facilitating the operation of language versions on websites

More than half of the websites available online use content management systems, i.e. CMS. They allow you to create and edit the content of the website without knowing any programming language or even HTML. This solution is very convenient and usually easy to use. CMS systems allow not only simple text editing, but are the basis for creating advanced web applications, such as online stores. According to data from <https://websitesetup.org/>, ten most popular CMSs used in 2019 are as follows¹:

- WordPress (27+ Million Live Websites)
- Wix (3.8+ Million Live Websites)
- Squarespace (1.9+ Million Live Websites)
- Joomla! Wix (3.8+ Million Live Websites)
- Shopify (1.1+ Million Live Websites)
- Drupal (630+ Thousand Live Websites)
- Blogger (430+ Thousand Live Websites)
- Prestashop (285+ Thousand Live Websites)
- Magento (265+ Thousand Live Websites)
- Bitrix (223+ Thousand Live Websites).

¹ <https://websitesetup.org/> (accessed on November 5, 2019).

Among the above-mentioned content management systems, the WordPress platform undoubtedly ranks first, with a market share of 51%. WordPress is used by over twenty-seven million websites². Out of the top 100,000 websites, 36.1% are based on WordPress³.

WordPress, as a platform for creating websites, is characterized by intuitive usage, a wide range of available themes, add-ons, plug-ins, easy installation, updating and expansion. Unlike other CMS systems, such as Drupal or Joomla, it does not require knowledge of PHP. All components are configurable from the Administration Panel through graphic forms. It is also possible to create and edit CSS and PHP code, also via the Administration Panel. WordPress is based on a free engine developed by developers working as volunteers. Anyone can report a bug or correct an existing one. The themes and plug-ins, on the other hand, are stored in a repository free of charge or under paid licences. Plug-ins usually have the support of developers, which allows for stable and safe operation.

WordPress can be used to create a website in any language or in several languages, however there is no built-in ability to create multilingual websites. There are powerful WordPress translation plug-ins that allow you to easily add multilingual content. The CMS itself can be installed in any language and the language of the Administration Panel can be adjusted for each user. With plug-ins, you can create bilingual or multilingual websites, you can even allow users to translate content via Google Translate. There are many types of translation plug-ins, but there are two basic types:

- Multilingual WordPress plug-ins
- Automated WordPress translation plug-ins.

Multilingual plug-ins allow you to add multilingual content to your website manually. The basic advantage of these plug-ins is that the quality of your translations will be much better than with any machine generated online translation tools.

The second type of WordPress translation plug-ins are those that use online translation services in order to translate the discussed content. These plug-ins do not require you to generate the content in multiple languages, although the translation quality is not as good as with manual translation.

² <https://trends.builtwith.com/> (accessed on November 6, 2019).

³ <https://websitesetup.org/> (accessed on November 5, 2019).

A list of the most popular plug-ins for language version management is as follows⁴:

- TranslatePress
- WPML
- Polylang
- Weglot
- Google Website Translator
- MultilingualPress
- TranslateWordPress with GTranslate
- Transposh WordPress Translation
- LocoTranslate.

TranslatePress is a WordPress translation plug-in that allows you to translate your WordPress website directly via the interface, with full support for WooCommerce, complex themes and website builders. The plug-in is classified within the group of automated translation plug-ins and its functioning consists in translating the entire page directly from the interface, with the support of Google Translate. After performing automatic translation, you can correct the text yourself. TranslatePress is released under the GPL licence and is self-hosted⁵.

WPML makes it an easy task to run a multilingual site on one of your WordPress instances. It is available in over 40 different language versions. You can also add your own language variants (e.g. Canadian French or Mexican Spanish) using the WPML language editor. It has an option of placing different language content in the same domain, in subdomains or in completely different domains. It can be easily integrated with the WooCommerce e-commerce plug-in (using WooCommerce Multilingual). It indicates which texts need to be translated and creates a complete and translated shop for the user⁶.

Polylang allows the user to create a bilingual or multilingual WordPress site. After the content has been added, you can define the language version for each of the entries. Translating a post, whether it is in the default language or not, is optional. You can use any number of languages. This plug-in belongs to a multilingual plug-in group, where the author provides the translation himself. RTL language scripts are supported. WordPress language packs are automatically downloaded and updated. The plug-in supports any type of content: posts, pages, media,

⁴ <https://www.wpbeginner.com/showcase/9-best-translation-plugins-for-wordpress-websites/> (accessed on October 24, 2019).

⁵ <https://translatepress.com/> (accessed on November 23, 2019).

⁶ <https://wpml.org/> (accessed on November 23, 2019).

categories, post tags, menus, widgets. Non-standard types of posts, non-standard taxonomies, sticky posts and post formats, RSS feeds, and all default WordPress widgets are supported. You can set the language by the content or language code in the URL, or you can use a different sub-domain or domain. Categories, post tags and some other meta are automatically copied when a new post is added or a page translated. You can use a professional or automatic translation service by installing LingotekTranslation as an add-on to Polylang. Lingotek offers a complete translation management system that provides services such as translation memory or semi-automatic translation processes. The Polylang plug-in has over half a million active installations⁷.

Weglot is another plug-in ensuring a reliable and easy way to translate your website content. Weglot is compatible with all CMS technologies; it automatically detects the entire content within the website in order to collect the text for translation. It is available in over 100 languages and has additional tools such as a contextual editor for translating the content directly into the website layout and structure. It has an option of machine translation via the first layer of automatic translation supplied by machine learning providers such as Microsoft, DeepL, Google and Yandex and it allows for the customization of translations⁸.

Google Website Translator is a plug-in that allows you to connect Google's automatic translation service to translate your website into over 100 languages. A simple and complete multilingual WordPress solution that can be used as a widget and as a short code. The plug-in is classified within the group of automated translation plug-ins⁹.

The second most popular website that offers both a content management system and a website builder is Wix, with a market share of 7.2%. Wix is used by over three million eight hundred thousand websites². Of the top 100,000 websites, 0.31% are based on Wix. This CRM offers users a free-of-charge and best website solution. It is a cloud-based development platform that allows users to create HTML5 websites and mobile websites via online drag and drop tools. As with WordPress, it is also possible to add social plug-ins used for e-commerce, online marketing, contact forms, email marketing and social forums. Wix website builder is based on freemium business model, achieving revenue through premium enhancements. The big advantage of Wix website builder is its drag-and-drop functionality for specific website elements. This allows you to connect with a corporate or personal site without large investments. The Administration Panel has

⁷ <https://wordpress.org/plugins/polylang/> (accessed on November 6, 2019).

⁸ <https://weglot.com/> (accessed on November 3, 2019).

⁹ <https://wordpress.org/plugins/google-website-translator/> (accessed on November 23, 2019).

a convenient and easy-to-use interface. Website optimization is done automatically and it is a more secure version than WordPress, mainly because the system engine is not made public.

The Wix system is paid and the starter plan forces ads to run on your website. Getting rid of the ads on your website is costly. As the Wix platform operates within a cloud-based system, the only option to sign up is to purchase an appropriate subscription plan, each of which is limited in functionality. This even applies to business and commercial plans sold at premium rates. The site builder itself also limits the options, as the only tools available are dragging and dropping items anywhere instead of closed boxes. It is fairly easy to add functionality to your Wix pages with the AppStore offered. The only problem is that most of the apps are not genuine apps. You can add frames instead of a genuine functionality to your website. The Wix platform has an SEO wizard that can help you optimize your content, but the techniques offered do not allow you to use the latest positioning and content search tools. The emphasis is mostly on keywords, grouping, and rankings regardless of other aspects of modern SEO¹⁰.

WixMultilingual is used for creating a multilingual website on the Wix platform. The process of adding more language versions can start after the entire website has been created. WixMultilingual creates different language versions of the same website (e.g. French version, Spanish version, etc.), and it does not create duplicate websites in every language version. The Wix system allows you to translate texts (in text fields, menu items, buttons, lists and grids, etc.), Wix forms, SEO settings, Wix chat, Wix store (some automatically generated store front texts do not appear in translation in all languages) and Wix events. Some text elements in the application have to be translated manually by the editor, the rest are translated automatically. Currently, the Wix platform does not have an option to translate content found in most business solutions, such as Wix Hotel, Wix Reservations, Wix Restaurants, etc. Similarly, the problem occurs with some applications: Wix Blog, Wix Forum, and third-party applications. It is also impossible to translate the member's zone, as well as the content added via Corvid database content on dynamic pages and older versions of some components¹¹.

Joomla! is a free and open source content management system for publishing web content. It is based on a web application structure with model and controller views that can be used independently of the CMS. With such a tool, it is possible to create powerful web applications. Joomla! provides a good basis for creating

¹⁰ <https://brandongaille.com/10-pros-and-cons-of-wix-websites> (accessed on November 22, 2019).

¹¹ <https://support.wix.com/en/article/creating-a-multilingual-site-with-wix-multilingual> (accessed on October 28, 2019).

your own modules and components, has built-in user support and an Administrator Panel. It is a good basis to be used by programmers who are experienced in programming in PHP and JavaScript. Joomla! is one of the most popular web programs with a global community of developers and volunteers who make the platform user-friendly, extendable, multilingual, responsive and optimized for search engines. Its market share is 3.45%. Joomla! is the basic system for over half a million sites. Of the top 100,000 websites, 0.99% are based on Joomla! It can be used to create corporate websites, portals, intranets and extranets. It is also a good tool for website preparation for small businesses, online magazines, newspapers and publications. It has plug-ins and modules allowing for the construction of e-commerce tools, online reservations, and social networks. Joomla! is highly extensible, with thousands of add-ons collected within the extension directory. If there are requirements that go beyond what is available in the base Joomla! suite, the Joomla application environment makes it easy for developers to create sophisticated add-ons that extend Joomla's capabilities in virtually unlimited directions. The core of Joomla! framework allows for creating various types of applications, such as asset control systems, reporting tools, custom product catalogues, integrated e-commerce systems, complex business catalogues, reservation systems, and communication tools. Joomla! is based on PHP and the MySQL relational database¹².

Joomla! allows you to easily create a multilingual website without having to install a third party extension. You just need to install another language from the Administration Panel. It will appear in the list of available language versions. After publishing it for your content, you just need to enable the Language Filter and Language Code plug-ins in Joomla! These built-in plug-ins will help Joomla! to filter the content according to the selected language version. Another element is adding content in the selected language version. Unlike WordPress, Joomla! does not have many add-ons and plug-ins that allow for dynamic creation of different language versions and for automatic translation of such versions¹³.

Language availability

In today's world of Internet society, with huge information repositories, the user cannot be restricted. In particular, the fact of not knowing the language should not block access to knowledge or data. Therefore, the general availability of website

¹² <https://www.joomla.org/about-joomla.html> (accessed on November 22, 2019).

¹³ https://docs.joomla.org/J3.x:Setup_a_Multilingual_Site/Creating_a_Content_Language (accessed on November 22, 2019).

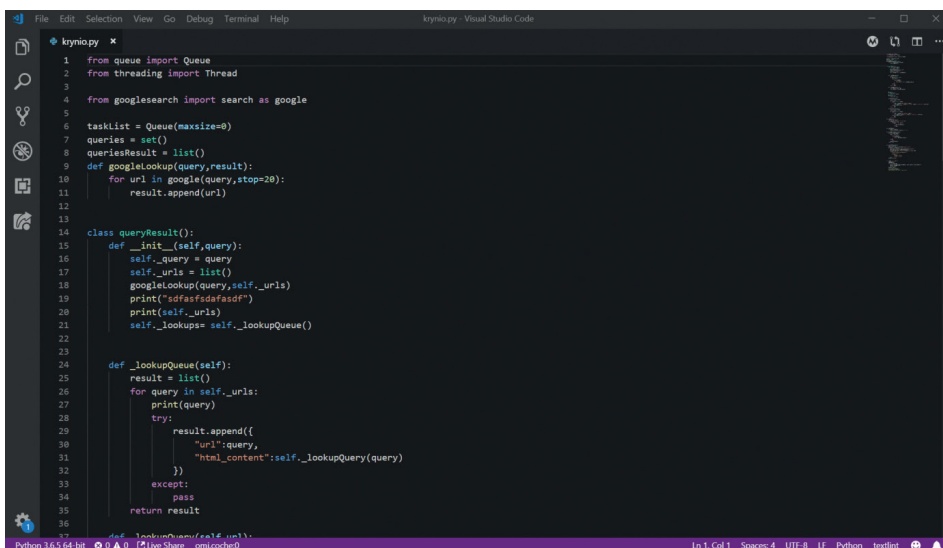
translation is already being implemented in the applications known as web browsers. An additional advantage is Google's introduction of an option of automatic real-time translation of the browsed pages.

The algorithms of automatic translation between different languages have been constantly improved for many years. Many years of experience have allowed for the improvement of the used neural networks, and even for the use of artificial intelligence. Google is distinguished primarily because of its access to a huge amount of websites and data on the Internet, but above all, by the entire mass of the Internet community, being able to interfere with the translated phrases and to improve them. They are not only professionals, linguists and software engineers, but also, and perhaps above all, ordinary Internet users who want to help by influencing the quality of translations.

What allows for effective translation? Why do some websites indexed in Google search still remain in the original language version? What is the cause of complications? How can you eliminate the inconvenience? Is it the manufacturing technology?

As there is a lot of web design technology, the focus was only on websites positioned within the Google search engine, websites where information material is presented within an editable system. Technologies such as Adobe Flash or Microsoft Silverlight have not been included in the analysis.

To automate the analysis process, rather than manually analyse hundreds of pages, a program has been written to achieve the results of the linguistic accessibility of websites. The program is based on the Python scripting language. Keywords – technical expressions – have been selected for analysis. Their availability and the results have been analysed in terms of language. Some of the search terms are: motion capture, artificial intelligence, machine learning, quantum computers, blue whale, t-rex, cisco, print cube, weed, snoop dog, wiz khalifa, kahoot, virtual reality, Augmented Reality.



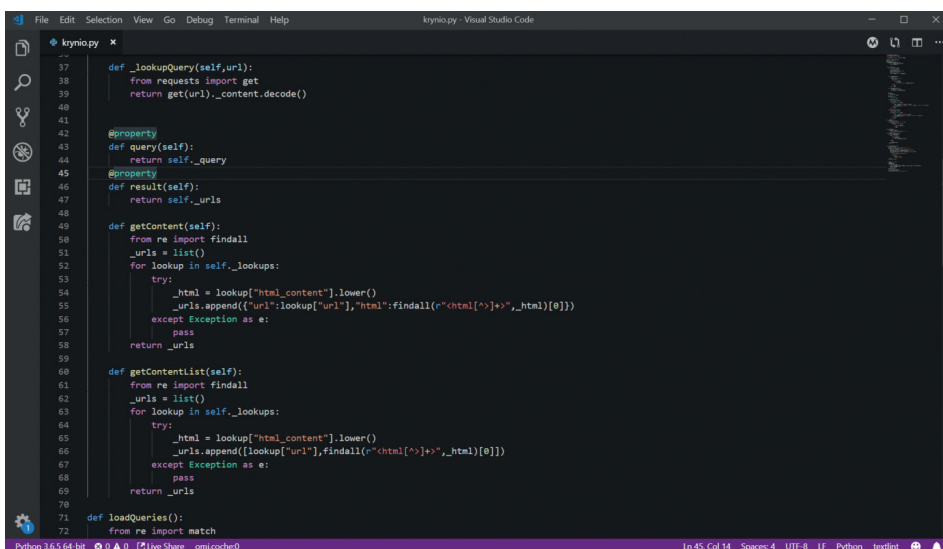
```

1 from queue import Queue
2 from threading import Thread
3
4 from googlesearch import search as google
5
6 tasklist = Queue(maxsize=0)
7 queries = set()
8 queriesResult = list()
9 def googlelookup(query,result):
10     for url in google(query,stop=20):
11         result.append(url)
12
13
14 class queryResult():
15     def __init__(self,query):
16         self_query = query
17         self_urls = list()
18         googlelookup(query,self_urls)
19         print("dfasfdasfdr")
20         print(self_urls)
21         self_lookups= self._lookupQueue()
22
23
24 def _lookupQueue(self):
25     result = list()
26     for query in self_urls:
27         print(query)
28         try:
29             result.append({
30                 "url":query,
31                 "html_content":self._lookupQuery(query)
32             })
33         except:
34             pass
35     return result
36
37 def _lookupQuery(self,url):

```

Figure 1. Defining search results in the headings of websites.

A file loaded into the script has been created from keywords, there, therefore, the analysis is universal. Words can be added and modified, and the script can be run several times. With ever-changing indexing of websites, the results are varied. This makes the analysis more complete.



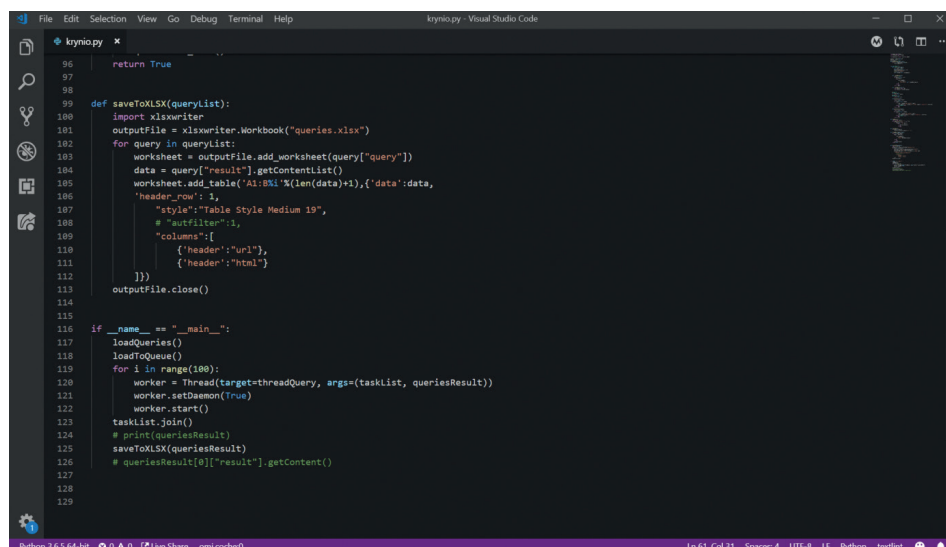
```

37 def _lookupQuery(self,url):
38     from requests import get
39     return get(url)._content.decode()
40
41
42 @property
43 def query(self):
44     return self_query
45
46 @property
47 def result(self):
48     return self_urls
49
50 def getContent(self):
51     from re import findall
52     _urls = list()
53     for lookup in self_lookups:
54         try:
55             _html = lookup["html_content"].lower()
56             _urls.append([lookup["url"],findall(r"<html[^\>]*>",_html)[0]])
57         except Exception as e:
58             pass
59     return _urls
60
61 def getContentList(self):
62     from re import findall
63     _urls = list()
64     for lookup in self_lookups:
65         try:
66             _html = lookup["html_content"].lower()
67             _urls.append([lookup["url"],findall(r"<html[^\>]*>",_html)[0]])
68         except Exception as e:
69             pass
70     return _urls
71
72 def loadQueries():
73     from re import match

```

Figure 2. Defining the content in the searched website headings.

In the next stage, the script automatically searches for the first twenty positioned pages, related to a given phrase in the Google search engine. It does this with every keyword defined in the file.



```

96     return True
97
98
99 def saveToXLSX(queryList):
100     import xlswriter
101     outputFile = xlswriter.Workbook("queries.xlsx")
102     for query in queryList:
103         worksheet = outputFile.add_worksheet(query["query"])
104         data = query["result"].getContentList()
105         worksheet.add_table("A1:B%i"%(len(data)+1),{"data":data,
106             "header_row": 1,
107             "style":"Table Style Medium 19",
108             # "autofilter":1,
109             "columns": [{"
110                 ("header":"url"),
111                 ("header":"html")
112             }])
113     outputFile.close()
114
115
116 if __name__ == "__main__":
117     loadQueries()
118     loadToQueue()
119     for i in range(100):
120         worker = Thread(target=threadQuery, args=(taskList, queriesResult))
121         worker.setDaemon(True)
122         worker.start()
123     taskList.join()
124     # print(queriesResult)
125     saveToXLSX(queriesResult)
126     # queriesResult[0]["result"].getContent()
127
128
129

```

Figure 3. Defining how data is to be collected.

The script returns the search results to a spreadsheet file where results can easily be analysed. The figures present examples of search results for a number of keywords in the area of technology.

As has already been mentioned, the website building techniques are different and therefore there is a need to find a common factor with which the obtained results can be analysed. It turns out that all techniques can be analysed with the use of the so-called website header, where the page parameters are defined. With such functionality, we can easily perform the analysis. For each page it finds, the script peaks the header and saves the gathered data in a file.

One of the basic parameters of the HTML language used to build web sites is the language parameter (LANG). By analysing its syntax and content, we can draw several conclusions.

Over a thousand websites have been analysed using twenty phrases – keywords. The examples of the results obtained for a number of selected phrases are presented in Figures 4, 5 and 6.

	A	B
1	url	html
2	https://en.wikipedia.org/wiki/Motion_capture	<html class="client-nojs" lang="en" dir="ltr">
3	https://en.wikipedia.org/wiki/Motion_capture#Disadvantages	<html class="client-nojs" lang="en" dir="ltr">
4	https://en.wikipedia.org/wiki/Motion_capture#Applications	<html class="client-nojs" lang="en" dir="ltr">
5	https://en.wikipedia.org/wiki/Motion_capture#Optical_systems	<html class="client-nojs" lang="en" dir="ltr">
6	https://en.wikipedia.org/wiki/Motion_capture#Non-optical_systems	<html class="client-nojs" lang="en" dir="ltr">
7	https://pl.wikipedia.org/wiki/Przechwytywanie_ruchu	<html class="client-nojs" lang="pl" dir="ltr">
8	https://www.gdys.sk/en/timetable/event/profi-workshop-2/	<html lang="en-us" prefix="og: http://ogp.me/ns#" class="html_boxed responsive av-preloader-disab
9	https://www.youtube.com/watch?v=kH7msPLVW_k	<html lang="pl-pl" data-cast-api-enabled="true">
10	https://www.youtube.com/watch?v=SN3Gj-XgPps	<html lang="pl-pl" data-cast-api-enabled="true">
11	https://www.xsens.com/motion-capture	<html class="no-js lt-ie9 lt-ie8 lt-ie7" lang="en" >
12	https://www.youtube.com/watch?v=fm-A1kmrxE	<html lang="pl-pl" data-cast-api-enabled="true">
13	https://www.vicon.com/about-us/what-is-motion-capture/	<html class="ie7 no-js" lang="en-us">
14	https://www.vicon.com/	<html class="ie7 no-js" lang="en-us">
15	https://www.optitrack.com/	<html lang="en" id="home" class="(none)">
16	https://www.qualisys.com/	<html class="no-js ie6 oldie" lang="en">
17	https://ar-tracking.com/applications/motion-capture/	<html dir="ltr" lang="en">
18	https://grendelgames.com/getting-started-with-motion-capture/	<html lang="en-us">
19	http://dashdot.pl/en/motion-capture/	<html lang="pl">
20	https://www.rokoko.com/	<html data-n-head="1">
21	https://www.engadget.com/2018/05/25/motion-capture-history-video	<html lang="en">

Figure 4. A spreadsheet presenting the results for the phrase: "Motion capture".

	A	B
1	https://en.wikipedia.org/wiki/Quantum_computing	<code><html class="client-nojs" lang="en" dir="ltr"></code>
2	https://en.wikipedia.org/wiki/Timeline_of_quantum_computing	<code><html class="client-nojs" lang="en" dir="ltr"></code>
3	https://en.wikipedia.org/wiki/Quantum_superposition	<code><html class="client-nojs" lang="en" dir="ltr"></code>
4	https://en.wikipedia.org/wiki/Quantum_supremacy	<code><html class="client-nojs" lang="en" dir="ltr"></code>
5	https://www.weforum.org/agenda/2019/10/quantum-computers-next-frontier-classical-google-ibm-nasa-supremacy	<code><html data-n-head-ssr lang="en" data-n-head="lang"></code>
6	https://www.ibm.com/quantum-computing/learn/what-is-quantum-computing/	<code><html lang="en-us" prefix="og: http://ogp.me/ns#"></code>
7	https://www.ibm.com/blogs/research/2017/09/qc-how-it-works/	<code><html data-n-head-ssr lang="en" data-n-head="lang"></code>
8	https://www.ibm.com/quantum-computing/learn/what-is-ibm-q/	<code><html lang="en-us" prefix="og: http://ogp.me/ns#"></code>
9	https://www.ibm.com/blogs/research/2017/09/quantum-molecule/	<code><html xml:lang="en-gb" lang="en-gb"></code>
10	https://www.sciencemag.org/quantum-computers	<code><html lang="en-us"></code>
11	https://www.newsscientist.com/article/2221107-timeline-a-brief-history-of-quantum-computing-from-1980-to-2010	<code><html lang="en-gb" prefix="og: http://ogp.me/ns# article: http://ogp.me/ns/article# video: http://ogp.me/ns/video#></code>
12	https://www.nytimes.com/2019/10/30/opinion/google-quantum-computer-sycamore.html	<code><html lang="en" class="grade-c"></code>
13	https://www.ft.com/content/b9b4e54d-dc1c-11e9-8f96-77216eb1f17	<code><html lang="en" class="grade-c"></code>
14	https://www.sciencemag.org/news/2019/10/ibm-casts-doubt-googles-claims-quantum-supremacy	<code><html data-reactroot="" data-reactid="1" data-react-checksum="1798475972"></code>
15	https://www.nature.com/articles/d41586-019-03213-z	<code><html lang="en" class="grade-c"></code>
16	https://www.quantamagazine.org/google-and-ibm-clash-over-quantum-supremacy-claim-2019-02-3/	<code><html lang="en" class="grade-c"></code>
17	https://www.nature.com/articles/d41586-019-02936-3	<code><html lang="en" class="grade-c"></code>
18	https://www.nature.com/2019/10/14/investment-in-quantum-computing-is-booming-but-will-a-quantum-winter-	<code><html class="ie8" lang="en"></code>
19	https://singularityhub.com/2019/10/14/investment-in-quantum-computing-is-booming-but-will-a-quantum-winter-	

Figure 5. A spreadsheet presenting the results for the phrase: "Quantum computing".

	A	B
1	https://www.sas.com/en_us/insights/analytics/machine-learning.html	<code><html class="sax" lang="en"></code>
2	https://en.wikipedia.org/wiki/Machine_learning	<code><html class="client-nojs" lang="en" dir="ltr"></code>
3	https://en.wikipedia.org/wiki/Boosting_(machine_learning)	<code><html class="client-nojs" lang="en" dir="ltr"></code>
4	https://en.wikipedia.org/wiki/Timeline_of_machine_learning	<code><html class="client-nojs" lang="en" dir="ltr"></code>
5	https://en.wikipedia.org/wiki/Deep_learning	<code><html class="client-nojs" lang="en" dir="ltr"></code>
6	https://www.coursera.org/learn/machine-learning	<code><html xmlns:fb="http://ogp.me/ns/fb#" itemtype="http://schema.org" lang="en"></code>
7	https://www.coursera.org/learn/machine-learning/reviews	<code><html xmlns:fb="http://ogp.me/ns/fb#" itemtype="http://schema.org" lang="en"></code>
8	https://www.coursera.org/learn/machine-learning/instructors	<code><html xmlns:fb="http://ogp.me/ns/fb#" itemtype="http://schema.org" lang="en"></code>
9	https://towardsdatascience.com/10-machine-learning-methods-that-every-data-scientist-should-know-3c96e0	<code><html lang="en"></code>
10	https://www.youtube.com/watch?v=kF9rgvfu	<code><html class="no-js gt-ie8" dir="ltr" lang="en" prefix="og: http://ogp.me/ns#" itemscope itemtype="http://schema.org/article"></code>
11	https://searchengine.pptel.techtarget.com/definition/machine-learning-ML	<code><html class="ie ie7" lang="en-us" prefix="og: http://ogp.me/ns#"></code>
12	https://www.geeksforgeeks.org/machine-learning/	<code><html lang="en"></code>
13	https://www.mathworks.com/discovery/machine-learning.html	<code><html class="no-js aws-lng-en-us" lang="en-us" data-static-assets="https://a0.awsstatic.com" data-js-version="1.0.308" data-js-version="1.0.308"></code>
14	https://aws.amazon.com/machine-learning/	<code><html lang="en"></code>
15	https://www.edx.org/course/machine-learning-with-python-from-linear-models-to-deep-learning-3	<code><html lang="en"></code>
16	https://www.edx.org/course/machine-learning	<code><html lang="en" class="grade-c"></code>
17	https://www.nature.com/articles/d41593-019-0520-2	<code><html lang="en" class="grade-c"></code>
18	https://medium.com/machine-learning-for-humans/whv-machine-learning-matters-6164ef1df12	<code><html lang="en"></code>
19		

Figure 6. A spreadsheet presenting the results for the phrase: "Machine learning".

Based on the research, we may conclude that the lack of a parameter identifying the website's language is rare. The graphical result of the analysis has been presented in Figure 7. LANG is the main parameter after which the Google search engine can suggest automatic translation.

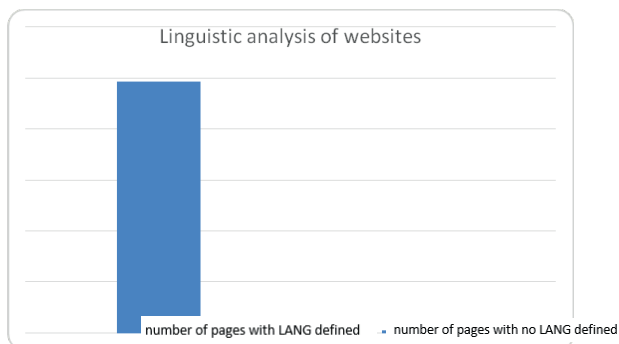


Figure 7. The result of the website coding declaration analysis.

Another conclusion that can be drawn basing on the conducted research is that the websites with a defined English language in the header, with content that can be analysed, i.e. in the text technique, have been proposed for translation. The websites that are created in the technology “unknown” to the search engine, i.e. where the text is in a non-editable form, are not offered for automatic translation. Automatic translation will also not be offered if the language of the website is not recognized or defined. The quantitative analysis of the language option in the website header has been presented in Figure 8.

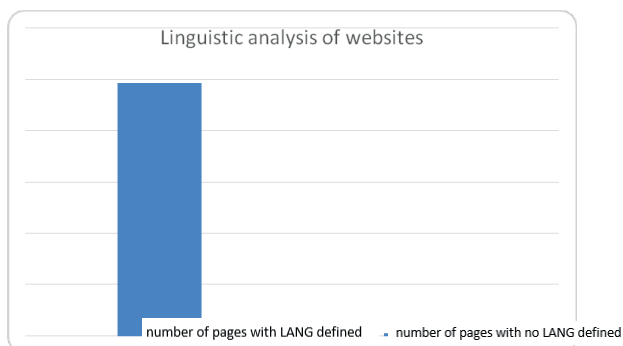


Figure 8. Quantitative analysis of the language option in the website header.

Conclusions

The access to knowledge is universal. The Internet should connect without any barriers. One of them, however, is language. Normally, we know the mother tongue and the one we learned during additional training and courses. Common technology works against people and allows texts in different languages to be understood more easily, thus reducing language barriers. What is more, it allows you to translate content or entire websites from one language to another, not just single words. This implementation was also placed in the Google search engine. With the HTML technology, it is possible to place a tag in the website header and define the language of the website. There is no need to translate the websites on your own, which is very convenient for Users.

With the variety of web design technologies, it would seem a challenge to define the language. It turns out that the core of each of the websites has a similar character, where you can place tags and keywords. The functionality results in a simple search for the desired content in any website, regardless of language or technological barriers.

References (online)

- <https://websitesetup.org/> (accessed on November 3, 2019)
- <https://trends.builtwith.com/> (accessed on November 6, 2019)
- <https://www.wpbeginner.com/showcase/9-best-translation-plugins-for-wordpress-websites/>
(accessed on October 24, 2019)
- <https://translatepress.com/> (accessed on November 23, 2019)
- <https://wpml.org/> (accessed on November 23, 2019)
- <https://wordpress.org/plugins/polylang/> (accessed on November 6, 2019)
- <https://weglot.com/> (accessed on November 3, 2019)
- <https://wordpress.org/plugins/google-website-translator/> (accessed on November 23, 2019)
- <https://brandongaille.com/10-pros-and-cons-of-wix-websites> (accessed on November 22, 2019)
- <https://support.wix.com/en/article/creating-a-multilingual-site-with-wix-multilingual>
(accessed on October 28, 2019)
- <https://www.joomla.org/about-joomla.html> (accessed on November 22, 2019)
- https://docs.joomla.org/J3.x:Setup_a_Multilingual_Site/Creating_a_Content_Language
(accessed on November 22, 2019)

Renata Uryga
ORCID No. 0000-0003-3346-332X
Andrzej Frycz Modrzewski Krakow University

Endangered information privacy

Abstract

In today's Internet world, our data is much more easily accessible to entrepreneurs, managers, etc. Never before have companies had so much public information. When segmenting consumer data (cookies), online habits can be read by looking at the pages they visit, the information they stop at, and the products they are looking for. It is easy to find us online, because we create websites ourselves, on which we place a lot of data about our private life. Thanks to this, companies are able to carefully select and adapt the message to a specific user so that it is as convincing as possible. All our movements leave a trace on the web, which are later used to define our behavior, likes, preferences, etc. This chapter presents the aspect of the possibility of collecting cookie information, as well as the Internet media used in marketing activities. Big data issues were described. It shows what information can be read from the visited pages and how effective data analysis influences decision-making support. The chapter also deals with the current state of website privacy protection.

Introduction

People have been analysing data for millennia. The invention of writing in ancient Mesopotamia was a tool used by bureaucrats to store and track information. The census in Biblical times was also used to gather information about citizens. Collecting and analysing such data in the analogue age was costly and time-consuming. The era of digitization has made great progress in data management, at the collection, storage and processing phases. Therefore, the big data phenomenon has emerged¹. Another term for the present day is danetization², which means collecting information about everything. The important thing about this concept is that “about everything” means collecting information about matters that we would never think of as sources of data. A report published on the EMC website³ in 2014 indicates that by 2020 the amount of data created and copied per year would reach 44 zettabytes, or 44 trillion gigabytes, as compared with 4.4 zettabytes in 2013. The digital age has made data processing faster and easier. Millions of calculations can be made within a second. Moreover, on this occasion it is worth asking a few questions. First, what is meant by information? Second, do we still have access to information, or have we already lost it? Third, is information privacy under threat? And fourthly, what do young people think about violating or not violating privacy? The purpose of the paper is also to present the attitudes presented by students of the Andrzej Frycz Modrzewski Cracow University regarding privacy in the context of data processing – especially in terms of the sense of privacy violation and the level of acceptance of this violation. Questionnaire studies have been used as research methods.

Data and information

Our life is surrounded by different data. As can be read in the report⁴, said data never sleeps. Data is collected from various sources. The data themselves do not provide information, but when combined with other data or subjected to some form of processing, they constitute certain information. Data are generated based on ad

¹ V. Mayer-Schonberger, K. Cukier, *Big data – efektywna analiza danych. Rewolucja, która zmieni nasze myślenie, pracę i życie*, trans. by M. Głatki, Warszawa 2017.

² *Danetyzacja*, <http://bc.wydawnictwo-tygiel.pl/public/assets/322/Cyber%20Media%2005.2019%20v01.04.pdf> (accessed on January 21, 2020).

³ Digital Report, <https://www.iotjournal.nl/wp-content/uploads/2017/01/idc-digital-universe-2014.pdf> (accessed on June 3, 2019).

⁴ Report, <https://www.domo.com/learn/data-never-sleeps-7> (accessed on January 20, 2020).

clicks, social media likes, transactions made, apps installed, streaming content and many more. From the Data NeverSleeps report, we can find out how much data were generated every minute of every day via the most popular platforms and companies in 2018 and 2019. Selected data have been summarized in the table below.

Table 1. The amount of data generated per minute.

Platform/Company	2019	2018
Email	188,000,000	103,000,000
SMS/MMS	18,100,000	12,986,111
YouTube	4,500,000	4,333,560
Google	4,497,420	3,877,140
Twitter	511,200	473,400
Instagram	55,140	49,380

Source: own study based on DOMO infographics.

According to a report published by DOMO, about 4,416,000 gigabytes of internet data are used in the United States in just sixty seconds. While that number was 41% lower in 2018, on average, in 2019 Google processed over 4,497,000 searches every minute. At the same time, we sent over 18,000,000 SMS messages globally. Nearly 188 million emails are being sent every minute. Within one minute, 511,200 tweets were posted on Twitter globally and 55,150 photos were posted on Instagram, 12 percent more than in 2018. In 2019, 4.5 million videos were viewed every minute on YouTube, and it is estimated that in 2019 there will be approximately 2.77 billion social media users globally, up from 2.46 billion in 2017. The Internet of Things is contributing to an ever-growing digital universe. In 2011,⁵ the number of IT devices exceeded the world's population. 20 billion of them were present in 2013. Forecasts indicated that by 2020 there would be 32 billion of such devices. Based on the statistical data, we can draw one basic conclusion. Collecting information about users has never been easier than today. Despite the enormity of the data leakage scandals that sparked a wave of public protest, people continue to voluntarily share information online. Only the most aware users have started to care more about their privacy, and we know that the more devices are connected

⁵ *Internet rzeczy*, <https://businessinsider.com.pl/technologie/internet-rzeczy-na-czym-polega-i-jaka-bedzie-jego-przyszlosc/Qqtr3t> (accessed on January 30, 2019).

to the Internet, the more we are monitored. And this amount of data will definitely make it difficult for our privacy to be protected.

Students' approach to privacy – survey

In order to explain the common understanding of the issue of privacy, the author conducted a survey among 245 students of the Andrzej Frycz Modrzewski Krakow University. The survey was conducted among students of various faculties (tourism, cosmetology, or management and IT), various levels of study (full-time, part-time, first and second cycle), as well as at different years. The group selected for the study were students, i.e. a group that should be well aware of the potential dangers of mass data processing and deal with information technology and the Internet actively.

During the preparation of the questions for the respondents, the focus was on issues related to modern technologies and new possibilities of processing large volumes of private data, including personal data. Two areas were taken into account. One of the important areas was the widespread use of the Internet, and the other area that was taken into account were the issues of mass processing of data originating from outside of the network.

The respondents answered the question: how, in your opinion, the collected data from various areas violates your sense of privacy, and the answers were on the scale from 1 – no sense of violation of privacy up to 5 – serious violation of privacy. The issues indicated in the survey included, among others various uses of the Internet. Issues related to the collection and gathering of information about the user behaviour on the Internet, gathering information based on tracking our activity on social networks, collection of information when using private file storage services in the cloud. The issues included mobile telephony, including breach of privacy through unauthorized access to text messages, MMS, etc. On the other hand, the issues raised in the survey included the processing of data originating from outside of the network, including, *inter alia*, collecting data with cameras – city and industrial monitoring, collecting data while using health services (e-sick-leaves, e-prescriptions), data collected from StreetView photos, satellite images, collecting data about payments (credit cards, payment cards), collecting data based on our location (GPS location), collecting data on devices referred to as the Internet of Things.

Data analysis

Table 2 shows the percentage results of the privacy violation part of the survey, according to the phenomena indicated.

Table 2. Questions and percentages of answers.

Issue	Reply				
	1	2	3	4	5
Common use of the Internet					
Use of websites	11%	9%	45%	22%	13%
Collecting files in the cloud	7%	16%	32%	22%	23%
Social media	5%	11%	31%	25%	28%
Access to SMS /MMS	9%	20%	32%	16%	23%
Banking cards	8%	16%	25%	23%	28%
GPS location	8%	12%	26%	24%	30%
Processing of data originating from outside the network					
Internet of things	21%	19%	32%	11%	17%
Data collection, e.g. from satellite images	14%	19%	32%	18%	17%
E-certificates, e-services	12%	24%	36%	16%	12%
City and industrial cameras	20%	15%	34%	11%	20%

Source: own analysis.

The survey indicates that the students were rather conservative in their approach towards the survey. Most often, they selected the middle answer. As for the widespread use of the Internet, the respondents realize that their privacy is being violated. On average, the level of privacy violation (with possible answers from 1 to 5) for individual questions varies and ranges from 3 to 5. Reading the results obtained within the second area, i.e. the processing of data originating from outside the network, it follows that this area does not concern the students directly and hence the level of privacy violation (with possible answers from 1 to 5) for individual questions is different and ranges from 1–3. In addition to the questions related to violating privacy, the author asked certain questions about the knowledge of the definition of big data or the Internet of Things. Most of the students (mainly those studying the humanities, 72% of the respondents) did not come across these

definitions before (about 56%). Additionally, the questionnaire included questions related to violation and protection of privacy. The question: are you aware that you are being followed online, received a positive reply from almost 57% of the respondents, and almost 27% replied that they are aware of it but did not care. There was also a question of violation of privacy when using the telephone. Almost 65% of respondents answered that they had the impression that their privacy had been violated. The students also answered a general question: Do you protect your privacy on the Internet. And here as many as 88% of the students answered that they protected their privacy. The issue of how students protect their privacy online will be the subject of further research. In addition, the results indicate that students of part-time and second-cycle studies, as well as students of science subjects, are more aware that they have to protect their privacy, and the questions about the level of the violation of their privacy were more often answered from 3 to 5. The surveys help to understand the understanding of the issue of privacy. They allow for the comparison of the level of violation of privacy related to various phenomena related to modern information technologies, including the processing of mass data using big data methods. At the same time, the research allows for dividing the population into groups according to, for example, the area of study or the mode of study.

Privacy policy

The “privacy policy” is defined as information posted on a website to inform users what personal data are collected therein and how they are used⁶. The data that the user will probably have to consent to include: access to the calendar, list of contact, history of viewed pages and bookmarks, access to applications active on a given device, access to the history of dialled numbers, access to all user profiles on a given device, access to content of SMS and email attachments, or even consent to the ability to change general phone settings. Panoptikon Foundation⁷ filed complaints to the President of the Office for Personal Data Protection against Google and IAB (Association of Internet Industry Employers). Panoptikon believes that the standards and the technical specifications for the auction of user data do not meet the requirements of GDPR. Most of the users use these devices almost around the clock, and the devices by default generate such metadata as: IP address,

⁶ *Polityka prywatności*, <https://poradnikprzedsiębiorcy.pl/-polityka-prywatnosci-w-sklepie-internetowym-wzor-z-omowieniem> (accessed on June 3, 2019).

⁷ Web tracking and profiling: *Ciemna strona śledzenia i profilowania*, https://panoptikon.org/sites/default/files/publikacje/panoptikon_raport_o_slledzeniu_final.pdf (accessed on June 3, 2019).

access time, session duration, type of software used, device location. Metadata – combined with information about how a specific person used an application or a service (where they clicked, what they looked for, what they purchased, how quickly they typed) – a detailed user profile is created, including their personality traits and a description of their habits and individual lifestyle. Such data are often sensitive and deeply interfere with our privacy. Dark patterns⁸ are tricks, “nasty practices” used on websites and apps that make us do things that we have not planned to do after much thought. This is the company’s deliberate activity, aimed at obtaining specific financial benefits, collecting data or obtaining consents for marketing mailing. We have to point out to the newsletters from which it is difficult to unsubscribe, or booking a flight, where all additional options are usually paid in advance and you should be careful when ordering a ticket.

Personalization or loss of access to information

In the 21st century “personalization”⁹ has taken on special importance. It has made it possible to offer a wide audience of various products and services directly tailored to the individual needs of each individual. By observing the dynamics of the progress of personalization activities, the user has the impression that they were created exclusively for him. Currently, personalization is taking all steps related to the adaptation of the content returned by search engines to the individual needs of each user based on the actions taken by him in the past in the virtual space. A user who, until now, has been browsing the Internet for the information on left-wing topics may never come across information on right-wing topics and vice versa, unless we browse the information consciously. And thus, something that was supposed to improve the method of obtaining information has restricted access to the information at the same time. Through activities undertaken on the Internet and data we share, each of us significantly narrows the space of information that we receive. It is a closed space that already gets filtered information, dedicated to individual users.

⁸ Dark patterns, <https://www.darkpatterns.org/> (accessed on January 21, 2020).

⁹ *Personalizacja*, https://ruj.uj.edu.pl/xmlui/bitstream/handle/item/72290/jablonska_personalizacja_internetu_zagrozenie_czy_naturalny_proces_2019.pdf?sequence=1&isAllowed=y (accessed on January 21, 2020).

Privacy threats

In May 2019¹⁰, news spread around the world about a gap in the popular WhatsApp application, with which hackers were able to install malware on their victims' phones. 0-day attacks rely on the use of previously unidentified vulnerabilities to break into the victim's system and remain a serious threat to corporate data. In March 2018¹¹, the media reported that in the "thisisyourdigitallife" quiz prepared by a Russian working at the University of Cambridge, Facebook obtained information from 87 million user accounts and also those who did not take part in this game. Norwegian Consumer Council¹² analysed the terms and conditions, including the privacy and behaviour policies of 20 mobile applications. The objective was to look for threats hidden in the end-user conditions and privacy in applications. With a huge number of applications (millions available on the market), NRK did not analyse the applications in depth, but its objective, above all, was to outline various problems. One of the problems noticed in the report¹³ was constant tracking of the users. The report indicated which techniques Google was using to make the user accept tracking. These techniques include manipulating click settings, hiding default settings, and misleading and unbalanced information. Another important issue is paying attention to the Internet of Things¹⁴ (IoT), which quickly spread over our personal and professional life. In everyday life, IoT solutions are widely used from smart cars, smart homes to a number of other small devices connected to the Internet and, which is important, connected with each other. All these devices process data about us. "Smart" technology is to make our lives easier, but it can also be used to interfere with our private life. We have an impression that they follow us and analyse our data. Household appliances that we use every day, such as refrigerators, dishwashers, televisions, cameras, child control devices, are connected to the Internet. Internet service providers can easily track our behaviour using IoT, even if these devices are used to protect our privacy. These devices have posed a threat to privacy since they were created, but now, due to the growing scale

¹⁰ *Cyberatak*, <https://www.cyberdefence24.pl/whatsapp-dotkniety-cyberatakiem-apel-do-wszystkich-uzytkownikow> (accessed on January 21, 2020).

¹¹ *Blog blisko technologii*, <https://www.spidersweb.pl/2018/12/facebook-podsumowanie-2018.html> (accessed on January 21, 2020).

¹² *I know your every step*, <https://fil.forbrukerradet.no/wp-content/uploads/2018/11/27-11-18-every-step-you-take.pdf> (accessed on January 21, 2020).

¹³ *Ibid.*

¹⁴ *Internet rzeczy*, https://www.sas.com/pl_pl/insights/internet-of-things.html?gclid=EAIaIQobChMI%20vbSqYYSK2%20wIVQ%20cAYCh0qZQQcEAYASAAEgJGwvD_Bw (accessed on December 30, 2019).

of their usage, said threat becomes even more important. Cybercriminals use IoT devices for their purposes, as exemplified by DDoS attacks¹⁵ (Distributed Denial of Service) or spying on people, e.g. by child monitoring devices. The size of data sets is growing rapidly, storage costs are decreasing, tools for analysing them are becoming more effective, and the security of our data storage and the protection of our privacy, despite various measures, is decreasing. At first sight, the functioning of the Internet of Things may look as if a dream came true for the world where burdensome duties are the domain of machines. But is it really so? We hear a lot about hacker attacks, cybercrime, cyberbullying, viruses and threats. And what if each of us is given power in the form of a telephone, a remote control with which we can do almost anything? Will we not become slaves of these little devices? What about online privacy? Online information on how much we earn, what our opinions are, what we read, who we are friends with, where we work and where we spend our nights? Even if we have not uploaded the information there ourselves? It is enough that the employer, a colleague with whom we were at the meeting, or the company where we won the competition for the best project, has uploaded such information. We do not have to express our opinion regarding a certain football team, but it is enough that we send an email to a colleague informing which of the teams we support. Internet remembers everything and basing on such data our personal profile is built. In the big data era, privacy has lost much of its effectiveness. The ability to capture personal data is built into the tools we use every day (browsers, smartphones, Internet of Things devices). There is no need to worry whether companies collect data to improve their results, but when they affect our private life or our devices, this is a cause for concern.

Internet user's profile based on data from various sources

The most popular data that we consciously place on the web are: name, surname, date of birth, address of residence, email address, username, passwords, interests, hobbies, declared profession, education or relationship status, searched queries. We put photos and information about the places we have been to. This information is relatively small. On the other hand, our online activity, applications used in the phones or Smart devices say much more about us. As we know, user behaviour patterns can provide a lot of information about the user. The algorithm from information collected from various data sources that we consciously put on the Internet,

¹⁵ DDoS attacks: *Cyberinstrumenty*, <https://repozytorium.uph.edu.pl/bitstream/handle/11331/.../Absi.J.Cyberinstrumenty.pdf> (accessed on January 21, 2020).

from active applications on smartphones, from household appliances that are connected to the Internet, to the phone and with each other will create a profile of our person. Such information will include: disciplines of sports we do, the number of people in the household, the diet we use, movies we like, language skills we possess, a house we have bought, etc. And the Internet, which promised complete freedom of choice, have been taking this freedom away, because when searching for information about an interesting place to stay with children, it provides us with a place in the mountains, since such terms have entered the search engine.

Conclusions

E-book readers, smart TV, thermostats, gas sensors, smart refrigerators, glasses, toothbrushes, toys, vacuum cleaners, smartphones are devices that supply databases with new information and indicators. They provide companies with uninterrupted access to information on any topic related to the interests and habits of customers. It is from these devices that the data are saved and often transferred in real time. In addition, a user voluntarily publishes information about their family, planned holidays or their interests online. The amount of the information appearing on the web is growing at an unimaginable pace. Even 20 years ago, those using search engines received identical results when typing the same keywords. It is practically impossible to find the desired search results freely. Despite many negative aspects of the use of information filtering mechanisms, personalization on the Internet is a natural process of web development. This development facilitates and improves communication and extracts interesting content, products or services on the Internet. However, the negative aspects that result from such personalization are serious and, as a result, may pose many threats. At this point, it should be emphasized that the scale of the threat is constantly determined by the activities undertaken by the Internet users. An Internet user can and should influence the reduction of the level of threat, and even its complete reduction.

We are aware that the state, corporations or companies analysing data know a lot about us and earn money on our weaknesses and influence our choices. We do not yet know what threat to the privacy of information may be posed by such data if they are used inappropriately and are not properly protected. Today, the Internet of Things has penetrated into almost every sphere of life. In the coming years, the trend towards smart devices will become so clear that it would be difficult to buy appliances or home appliances that are not connected to the Internet. We are so delighted with easy access to information caused by digital civilization that we forget about loss of access to information and loss of privacy. We cannot distance

ourselves from the digital world, our digital profile. It is not possible for us to remove it. If we really want to protect ourselves against the loss of privacy, then the only way is to avoid every electronic device.

References

Mayer-Schonberger V., Cukier K., *Big data – efektywna analiza danych. Rewolucja, która zmieni nasze myślenie, pracę i życie*, trans. by M. Głątki, Warszawa 2017

References (online)

Cyberinstrumenty, <https://repozytorium.uph.edu.pl/bitstream/handle/11331/.../Absi.J.Cybereinstrumenty.pdf> (accessed on January 21, 2020)

Blog blisko technologii, <https://www.spidersweb.pl/2018/12/facebook-podsumowanie-2018.html> (accessed on January 21, 2020)

Ciemna strona śledzenia i profilowania, https://panoptykon.org/sites/default/files/publikacje/panoptykon_raport_o sledzeniu_final.pdf (accessed on June 3, 2019)

Cyberatak, <https://www.cyberdefence24.pl/whatsapp-dotkniety-cyberatakiem-apel-do-wszystkich-uzytkownikow> (accessed on January 21, 2020)

Danetyzacja, <http://bc.wydawnictwo-tygiel.pl/public/assets/322/Cyber%20Media%2005.2019%20v01.04.pdf> (accessed on January 21, 2020)

Dark patterns <https://www.darkpatterns.org/> (accessed on January 21, 2020)

Digital Report, <https://www.iotjournaal.nl/wp-content/uploads/2017/01/idc-digital-universe-2014.pdf> (accessed on June 3, 2019)

I know your every step, <https://fil.forbrukerradet.no/wp-content/uploads/2018/11/27-11-18-every-step-you-take.pdf> (accessed on January 21, 2020)

Internet rzeczy, https://www.sas.com/pl_pl/insights/internet-of-things.html?gclid=EAIaIaQobChMI%20vbSqYYSK2%20wIVQ%20cAYCh0qZQQcEAAYASAAEGJGwvD_Bw (accessed on December 30, 2019)

Internet rzeczy, <https://businessinsider.com.pl/technologie/internet-rzeczy-na-czym-polega-i-jaka-bedzie-jego-przyszlosc/0qtrr3t> (accessed on January 30, 2019)

Personalizacja, https://ruj.uj.edu.pl/xmlui/bitstream/handle/item/72290/jablonska_personalizacja_internetu_zagrozenie_czy_naturalny_proces_2019.pdf?sequence=1&isAllowed=y (accessed on January 21, 2020)

Polityka prywatności, <https://poradnikprzedsiębiorcy.pl/-polityka-prywatnosci-w-sklepie-internetowym-wzor-z-omowieniem> (accessed on June 3, 2019)

Report, <https://www.domo.com/learn/data-never-sleeps-7> (accessed on January 20, 2020)

Marian Krupa

ORCID No. 0000-0003-3346-332X

Cavalry Captain Witold Pilecki State University of Małopolska in Oświęcim

Maciej Pękała

ORCID No. 0000-0002-4785-3583

Andrzej Frycz Modrzewski Krakow University

The use of “Check Alert” technique in data quality management and business analysis within the company

Abstract

The value of business analysis is directly related to the quality of data acquired to their preparation. The Business Data Management (BDM) process is extremely difficult both from internal and external perspective and it is exposed to different type of risk. The objective of this project is to evaluate the usefulness of the authorial “Check Alert” method in area of BI reports reliability verification performed by key users / business data owners. The paper consists of: Business

Data Quality Management accordingly to 8R principle; Business Data Management process and Risk Management; Characteristics of Simulation Modelling method; The “Check Alert” and SAP Analytics Cloud software implementation project – case study. The last part of the paper presents final conclusions and recommendations.

Introduction

The usefulness of business analyses is directly proportional to the quality of the data used to prepare the discussed analyses. The process of managing business data, both internally and externally, is extremely difficult and involves various types of risks. It usually proceeds in a standardized manner, within the framework of pre-defined and approved procedures or system algorithms operating within Business Intelligence software. However, an important problem is not the mere fact of efficient delivery of the report in an appropriate manner and within the required time, but also the reliability of the results of such a report, which have an extremely significant impact on the quality of decisions made within the company. Like any other process, the data (business information) management process also requires the use of appropriate methods and tools. The effectiveness in ensuring high quality data through planned and implemented system solutions finally determines the quality of the entire information management system in every company.

The objective of this paper is to assess the usefulness of our own “Check Alert” technique in supporting the key user/data owner in the process of verifying the reliability of the analysed reports, generated in BI class systems. The paper is an attempt to answer the following questions: 1) How is the data quality management process carried out in the company according to the 8R and 5 SDR principles? 2) What are the so-called “Data defects” of business information stored in the databases? 3) What is simulation modelling and why is it useful in improving the business data management process? 4) What is the proprietary diagnostic technique “Check Alert”?, and 5) How to use SAP Analytics Cloud software in the qualitative assessment of business data from the perspective of assessing the reliability of the generated reports?

In order to answer the above questions, the analytical and the synthetic method applied in scientific and industry-related literature was used, together with simulation modelling carried out using BI SAP Analytics Cloud software and the “Check Alert” technique.

The summary of the analysis and simulation, based on the financial data of the listed companies, becomes a summary of conclusions and recommendations prepared in the form of synthesis.

1. A company's business information management system in the company in accordance with the 8R and 5 SDR principle – towards the integration of logistic and information flows

The issues of data quality management, including business analyses, are discussed in the subject-matter literature, along with the context of analysing information disability ("information disease"). The above pathological behaviours have been presented from the point of view of both: the product as an effect of the information and communication management process implemented in the company, and the process that leads to the production of the above-mentioned products, such as reports and business analyses¹. Data malfunction with regard to information as product includes the following²:

- Information overloads – providing an excess of information and business data, which is manifested in an extended analysis time, higher costs, risk of inconsistency and lower motivation and commitment of a decision-maker;
- Information ambiguity – development of a report, analysis that leads to contradictory, ambiguous conclusions;
- Information anaemia or weakening of the "field of view" – a set of incomplete data, out of date data or an excess of data inadequate to the decision problem solved;
- Information distortion – incorrect association of data categories with their content, or two different contents or data sets are provided for the same category;
- Incorrectly structured forms as carriers of unreadable information – complicated structures, no logical connection of individual tables and fields, excess of manually entered text and data, chaotic editorial work.

As regards the qualitative assessment of the information management process, particular attention should be paid to the following³:

- Backlog of information – the so-called "Bottlenecks" in the process of creating and transferring data, business analysis;
- Information "distortion" – the ambiguity of information and data resulting from various formulas, standards for its input, processing and analysis;

¹ Z. Martyniak, *Podstawy diagnozowania informacji jako produktu i procesu*, [in:] Z. Martyniak (ed.), *Zarządzanie informacją i komunikacją. Zagadnienia wybrane w świetle studiów i badań empirycznych*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 2000, p. 11.

² Ibid., pp. 12–17.

³ Ibid., pp. 18–20.

- Information collapse – lack of data, lack of information in the system caused by lack of required resources, efficient procedures or lack of integration of goods and financial flows with information flows.

An extremely important element of preventing the so-called “Information failures” in the information management process is to ensure proper integration. Both in commodity flows (logistic perspective) and in information flows (decision perspective) the basic determinant of success is the implementation of an integrated system, for example based on the concept of eight conditions, requirements (8R) and 5 principles of rational management (5 SDR).

In the first case, the 8R concept (*8 Right*) indicates a need to provide the following information and communication management system in the company: 1) appropriate information (*right information*); 2) in the right amount (*right quantity*); 3) of adequate quality (*right quality*); 4) to the right place (*right place*); 5) in good time (*right time*); 6) to the right recipient, decision maker, employee (*right customer*); 7) at an appropriate level of costs (*right cost*) and 8) in a responsible manner from the point of view of the social and natural environment (*right way*)⁴. The 8R concept defines the necessity to simultaneously monitor and assess the quality of data and business analysis system from the perspective of the above-described perspectives. It is a relatively complex task requiring ensuring appropriate resources and methods, including advanced information technologies.

On the other hand, according to the 5 SDR, an integrated information management process based on the highest quality standards (8R principle) should lead to decision making in order to: 1) ensure a proper balance between economic goals and social/ecological costs (optimization principle); 2) eliminating contradictions between various decision-making aspects (the principle of cooperation); 3) minimizing contradictions, avoiding conflicts of goals (the principle of compatibility); 4) making decisions based on verified and reliable data (the principle of credibility); 5) taking into account industry, technological, legal, etc. conditions (the principle of competition)⁵.

Both the 8R concept and the 5 SDR model set us the fundamental premises leading to the implementation and maintenance of high quality information system management. In the further part of this study, the aspect of the business

⁴ J. Mangan, Ch. Lalwani, T. Butcher, *Global Logistics and Supply Chain Management*, Wiley & Sons, New York–London 2008.

⁵ A. Wiktorowska-Jasik, *Organizacja procesu transportu ładunków transportem zintegrowanym*, [in:] I.N. Semonov (ed.), *Zintegrowane łańcuchy transportowe*, Difin, Warszawa 2008, pp. 40–41.

decision-making process in the company will be described in more detail in the perspective of the so-called "Data defects" in database administration.

2. "Data defects" and the quality of business decisions made in the company

In general, the data that make up the information management system in the company are stored in databases. They are related to such data properties as: sharing, integration, integrity, security, abstraction and independence⁶. In our analysis, we will be interested in data integrity, i.e. a precise reflection of the area of analysis, the base of which is a model. In practice, this means that if there are real-world relationships between the objects represented by the data in the database, then the changes made in fact must be reflected in the database, and changes in one part of such a relationship in the database should be accurately reflected in the other part of the relationship. Integrity means that we should have confidence in what is in the database regarding the correspondence between the facts stored in the database and reality.

This trust can be undermined by the so-called "Data defects" understood broadly, in particular as damage, defect or missing. The origins of such defects may be different, and indirectly it is a classification (categorization) of defects, as shown in Figure 1⁷.

⁶ A detailed discussion of these characteristics can be found in P. Beynon-Davies, *Database systems*.

⁷ D. Gałęzowski, M. Sienkiewicz, *Zarządzanie jakością danych*, presentation in Warsaw SGH 06/12/2016, <https://studylibpl.com/doc/684763/jako%C5%9B%C4%87-danych> (accessed on July 27, 2019).

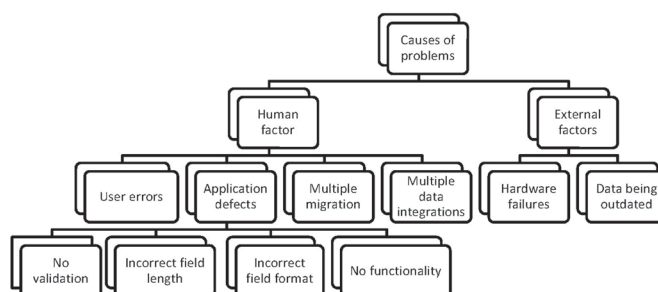


Figure 1. Origins of defects. Prepared on the basis of the presentation by the Warsaw School of Economics.

Source: D. Gałęzowski, M. Sienkiewicz, *Zarządzanie jakością danych*.

Janicki⁸ thinks that the quality of database is dependent on the completeness of information, correctness of information, consistency of data, timeliness and domain. Let us briefly introduce the above concepts:

- completeness of information – the degree of filling the database records with data;
- correctness of information – means their compliance with the actual state of affairs;
- data consistency – means correctness of mutual relationships, correctness of connections between the parameters, non-exclusion of data and references;
- topicality – is compliance with the facts despite the constant changes of reality;
- domain – is an area of information to be collected.

At the same time, consideration should be given to the causes of poor data quality. Janicki⁹ lists the following:

- lack of completeness at the uploading stage;
- lack of correctness at the uploading stage;
- no updates;
- field mismatch;
- the evolution of IT systems;
- conversions between IT systems;
- extending the use of information contained in the database;
- processing errors.

⁸ W. Janicki, *Jakość bazy danych*, "Automatyka", 9 (2005), item 5; <http://journals.bg.agh.edu.pl> (accessed on July 27, 2019). Similar features can be found in: K. Błaszczuk, R. Knosala, *Problem jakości danych w hurtowniach*, "Prace Naukowe Akademii Ekonomicznej w Katowicach" (2006), <http://www.swo.ae.katowice.pl> (accessed on July 27, 2019).

⁹ W. Janicki, *Jakość bazy danych*.

They fall into the categories of defects listed in Figure 1. There is an obvious question about the possibility of ensuring good data quality and effective defect detection. In general, these tasks fall within the scope of the activities carried out by the IT department or a specialized unit responsible for data quality. Can, however, a user-recipient of data be confident about their correctness and reliability despite the developed systemic solutions? Janicki¹⁰ notes that poor data quality may affect, *inter alia*, the following areas¹¹:

- customer identification and service;
- data processing;
- preparation of reporting data;
- management data;
- marketing activities.

In these areas, the data is used by a user to make various decisions. In general, incorrect data result in wrong decisions. Is the decision-maker only reliant on accepting the received information or reports as error-free? Perhaps there is a way of detecting errors, "data defects" based only on the analysis of reports or charts?

Based on the simulation modelling method and the proprietary "Check Alert" (CA) technique, an attempt was made to develop a systemic solution used just to support the data owner in the process of verifying the reliability of the analysed reports. Ultimately, it is to allow a recipient of reports from the BI class system to indicate potential ranges of source data that may be considered to be incorrect.

3. Simulation modelling – description of the research method

Simulation modelling is an extremely useful method of verifying hypotheses defined in multiple academic projects, including research carried out in the field of organizational and management sciences. In this paper, simulation, simulation modelling are understood as "imitation, imitation of a real system by means of experiments carried out on a model representing this system"¹². An indispensable element allowing for carrying out research and experiments under the above

¹⁰ Ibid.

¹¹ In practice, this leads to the following defects at the table, data cell level: 1) value omission (blank field); 2) multiple use of the same value; 3) omitting or adding one or more digits, and 4) changing the order of digits within the same value.

¹² M. Beaverstock, A. Greenwood, E. Lavery, E. Nordgren, *Applied simulation. Modelowanie i analiza przy wykorzystaniu FlexSim*, FlexSim Software Products, Orem, Utah, 2011, p. 6.

method is the need to develop and define an appropriate model – the aspect of verification and validation and the selection of an appropriate IT system¹³.

There are, with no doubt, numerous advantages of this method, both in terms of research and managerial practice. These include: big realism of the model and data, the ability to diagnose solutions/strategies that do not exist to date, the ability to take into account the time/seasonality parameter, flexible selection of optimization criteria/priorities/business goals, automation in the field of testing many different variants/business scenarios, minimization of knowledge and skills mathematics, low cost and relatively short time to acquire the required business knowledge. The disadvantages of this method would be as follows: the inability to draw general conclusions and recommendations that go beyond the scope and subject of the research, the need to carry out many experiments – time conditions, problems with indicating optimal solutions – the issue of selecting priorities, risks related to data manipulation in order to obtain the expected results or also the verification of previously adopted hypotheses¹⁴.

In this study, as part of simulation modelling, the author's own "Check Alert" technique was developed, which allows, using BI class systems, to make a qualitative assessment of the processed data and try to indicate the sources of the so-called "Data defects".

4. The use of the "Check Alert" technique and SAP Analytics Cloud software in the qualitative assessment of business data – case studies

The "Check Alert" technique consists in carefully reviewing the graph or charts reflecting the processed and analysed data and paying attention to the detected irregularities. These may in particular include the following:

1. peaks, i.e. values representing global extremes, i.e. the largest or smallest values in the data set clearly lying outside the plot line;
2. flattening, i.e. several identical values occurring next to each other;

¹³ A. Balcerak, W. Kwaśnicki, *Modelowanie symulacyjne systemów społeczno-gospodarczych: różnorodność podejść i problemów*, [in:] A. Balcerak, W. Kwaśnicki (eds.), *Symulacja systemów społeczno-gospodarczych*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2005, pp. 6–7: "Simulation should be considered a specific technique of studying reality, developed particularly intensively in recent decades. The development of this technique was possible thanks to revolutionary changes in computer technology (...), therefore, when speaking about simulation today, we usually mean computer simulation".

¹⁴ Ibid., p. 6.

3. contradiction with experience, i.e., noticing that the appearance of the chart does not corroborate the user's expectations resulting from experience and knowledge;
4. an additional warning signal for the user may be the behaviour of the plotted trend line.

The "Check Alert" technique does not automatically indicate a data defect, because in fact there are unexpected situations causing anomalies on the chart. Examples include: sharp rises or falls in stock prices after information about the destruction of factories producing components or mines extracting strategic resources as a result of a disaster (e.g. tsunami); or information about military coups in countries with a dominant role in supplying some raw material causes a peak on the chart of changes in stock prices; terrorist attacks on strategic objects or infrastructure. The final decision whether an irregularity detected by the user is the result of a data defect or results from unexpected changes in the economic reality in the external or internal environment must be made through the implementation of an appropriate procedure with the participation of the data quality section from the IT department.

In the analysis presented in this paper, the "Check Alert" technique has been used in order to analyse simulations of events in three different companies (Enea S.A., Orbis S.A. and Quantum Software S.A.). For the purposes of the simulation, the original data have been modified to include "data defects"¹⁵ introduced deliberately. Then, reports were prepared using the advanced Business Intelligence system (SAP Analytics Cloud) and the reports contained charts with the trend line and a forecast for the two future quarters¹⁶. In each of the cases discussed in the paper, graphs have been provided and the type of detected anomalies has been indicated with the potential source of the anomaly (the source of the defect).

In point 3, the definition of simulation refers to the model of the system, while the experiments described below have been performed using the commercial

¹⁵ In order to perform the simulation, the following changes (defects) in the data tables were made in a manner unknown to the tester-analyst: 1) For Enea SA, a) for the "Sales revenues" feature, the values were repeated in the field 2016Q2 from 2016Q3; b) for the 'Equity' feature, in field 2018Q2, the value 13 was changed to 31; c) for the "Book value of 1 share", "1" was added in the 2018Q4 field, changing the value from 39.93 to 311.93; 2) For the "Orbis S.A" company, for the "Equity" feature for the 2018Q4 field, the value was changed by removing 1 digit, i.e. from a value of 2,386,425 to 386,425; 3) For Enea S.A., for the feature "Sales revenues" while preparing the appropriate structure of the data file, the data from the "Book value of 1 share" feature was copied and thus the data was repeated, two features had identical data.

¹⁶ All possible trend lines available in the system were used to illustrate the program's functional capabilities. However, it should be remembered that there is no universal method of forecasting using the trend.

system version provided by the manufacturer for educational purposes and the actual data modified for the experiment.

4.1 Enea S.A. – the first data set

Using the SAC program¹⁷, forecasts were made taking into account three available methods for three data groups: equity, sales forecast, book value of 1 share. The next three figures indicate the effects of the analysis and forecast.

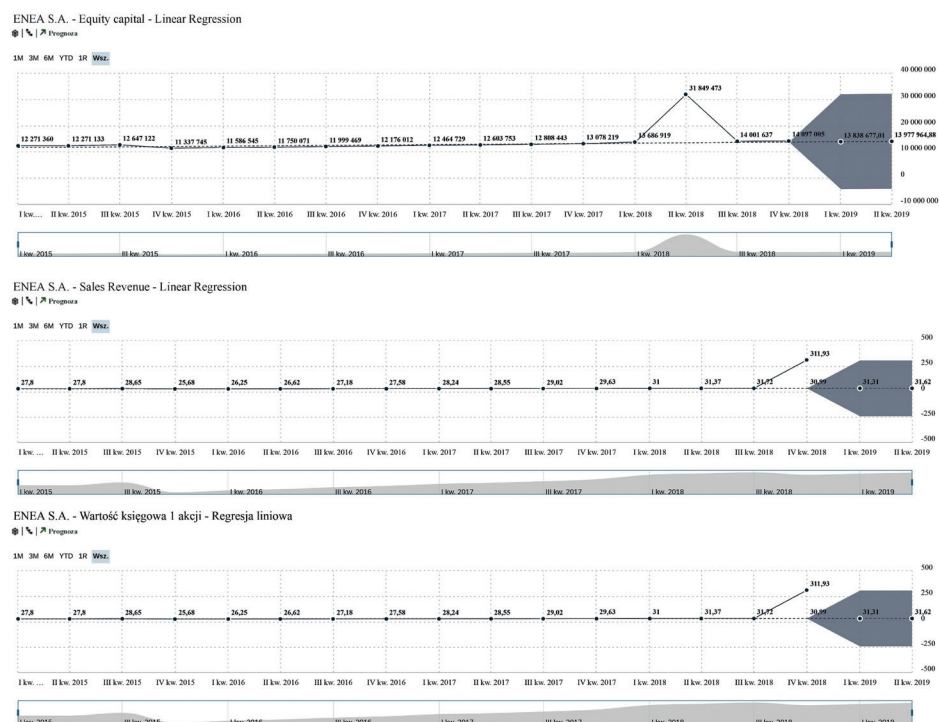


Figure 2. Enea S.A., method: linear regression.

Source: SAC system report.

¹⁷ SAC was used in the *SAP Analytics Cloud for Higher Education version* with the “Academic Account” license. It includes a complete business analysis, planning, predictive analysis and the *SAP Digital Boardroom* functionality within a shared learning environment.

ENEA S.A. - Equity capital - Triple exponential smoothing

● | % | P Prognosis



ENEA S.A. - Sales Revenue - Triple exponential smoothing

● | % | P Prognosis



ENEA S.A. - Book value of 1 share - Triple exponential smoothing

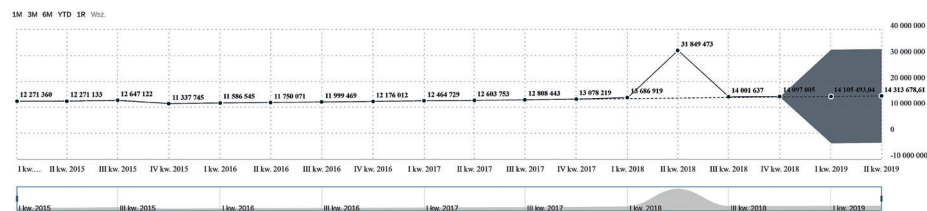
● | % | P Prognosis



Figure 3. Enea S.A., method: triple exponential smoothing.

Source: SAC system report.

ENEA S.A. - Equity capital - Automatic planning



ENEA S.A. - Sales Revenue - Automatic planning



ENEA S.A. - Book value of 1 share - Automatic planning

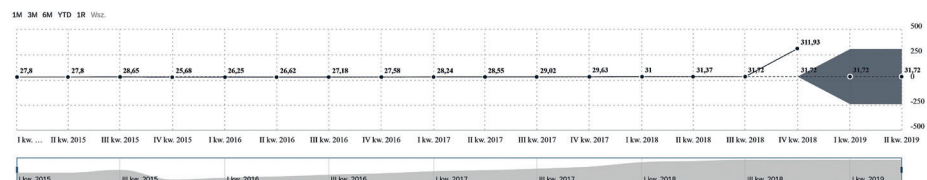


Figure 4. Enea S.A. method: automatic planning.

Source: SAC system report.

The recipient (data user) who analyses the above charts should notice three things.

1. Looking at the equity chart, the first in each of Figures 2–4, we can see a clear jump in value in Q2 2018. The dashed line, the chart of values calculated according to the selected forecast method, suggests a much lower value in the value line of the source data. The sharp jump in capital and an equally sharp decline during the quarter suggest a data error.
2. Similarly, in the graph of sales revenues, the second in each of Figures 2–4, we have a clear increase in the value of sales in Q4 of 2018. This, in turn, results in an optimistic increase in sales in the first half of 2019. Here, it is not obvious that the increase could not have happened, but it is too big even for the fall and winter months. The value of this quarter, calculated according to the selected method of forecast is placed in the line of the value of the source data, which suggests that the value should be lower.
3. The charts for sales revenue and the book value of 1 share are identical, which means that one of the charts had an incorrectly assigned data source. By analysing

the data values, the user may expect that the book values of 1 share have been assigned to the sales revenues.

4. All cases are the result of potential data defects and should be reported for clarification in the data quality assurance section (to the IT department). Upon obtaining an acknowledgement of the removal of defects, the user has to "refresh" the charts – repeat the forecast calculations, in accordance with the applicable procedure. The recipient (user) was informed that the data defects had been removed and, after running the report, they obtained the charts as shown in Figures 5–7.

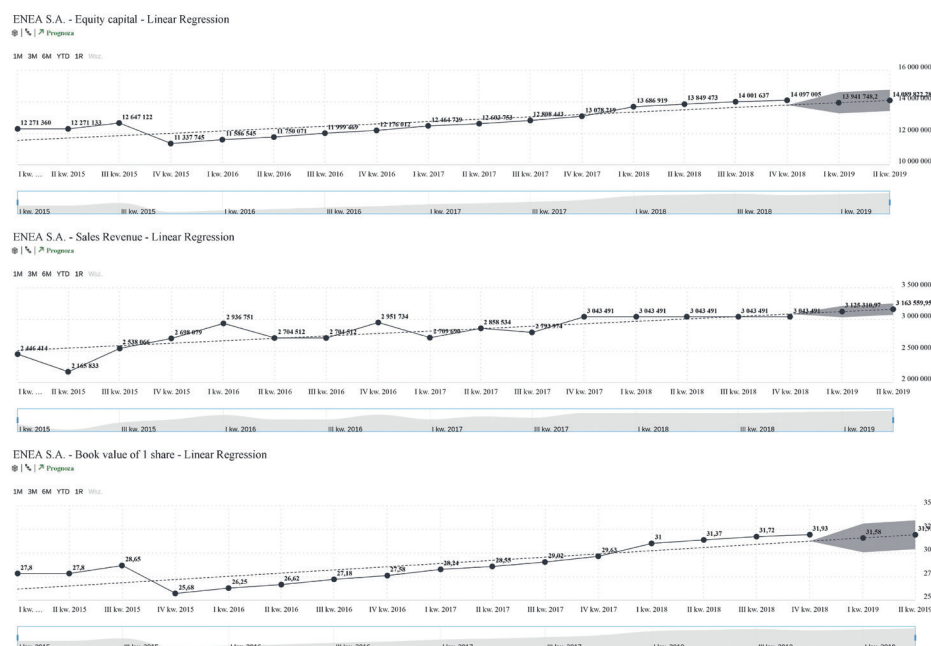


Figure 5. Enea S.A., method: linear regression.

Source: SAC system report.

ENEA S.A. - Equity capital - Triple exponential smoothing

• | % | Prognosis

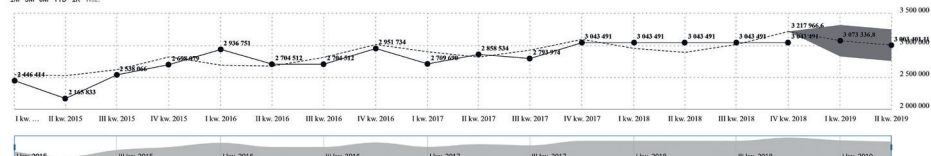
3M 3M 6M YTD 1R WnZ



ENEA S.A. - Sales Revenue - Triple exponential smoothing

• | % | Prognosis

3M 3M 6M YTD 1R WnZ



ENEA S.A. - Book value of 1 share - Triple exponential smoothing

• | %

3M 3M 6M YTD 1R WnZ



Figure 6. Enea S.A., method: triple exponential smoothing.

Source: SAC system report.

- ¹⁸ W. Janicki, *Jakość bazy danych*.

- conversions between IT systems;
- processing errors.

Removing defects and preventing them in the future is the duty of the data quality assurance section (IT department).

4.2 Orbis S.A. — the second data set

The second example is analogous to the defect considered in subchapter 4.1 for the book value of 1 share, the data lies outside the data range, which will appear as a peak in the chart. Using SAC, the data was analysed, this time only one method of calculating the trend was used and the results were obtained as in Figure 8.

Equity capital - Automatic planning



Figure 8. Orbis S.A. — equity, method: automatic planning.

Source: SAC system report.

The last data for Q4 2018 clearly differs from the line of the remaining data. The situation is similar as in the graphs in Figure 4, except that the data for the last quarter of 2018 are clearly below the graph line. Verification of the value of equity capital for Q4 2019 confirmed the existence of an error, and after correction, we received the same results as in Figure 9.

Equity capital - Automatic planning

Orbis SA

📊 | 📅 | 📈 Progress

1M 2M 6M YTD 1R Wtd:



Figure 9. Orbis S.A. — equity, method: automatic planning, data after adjustment.

Source: SAC system report.

Such an evident case is rare, but it can be a hint for verification of data entered “manually”, i.e. a defect within the category “user errors” (Figure 1) or “incorrectness at the input stage”¹⁹. With data from the database, this should not happen, nevertheless, reading the database from incorrectly operating devices may also be the cause of a similar error.

4.3 Quantum Software S.A. — the third data set

Now consider an example that may also contain erroneous data, i.e. the existence of a data defect. We used SAC again and got the same result as in Figure 10.

¹⁹ Ibid.

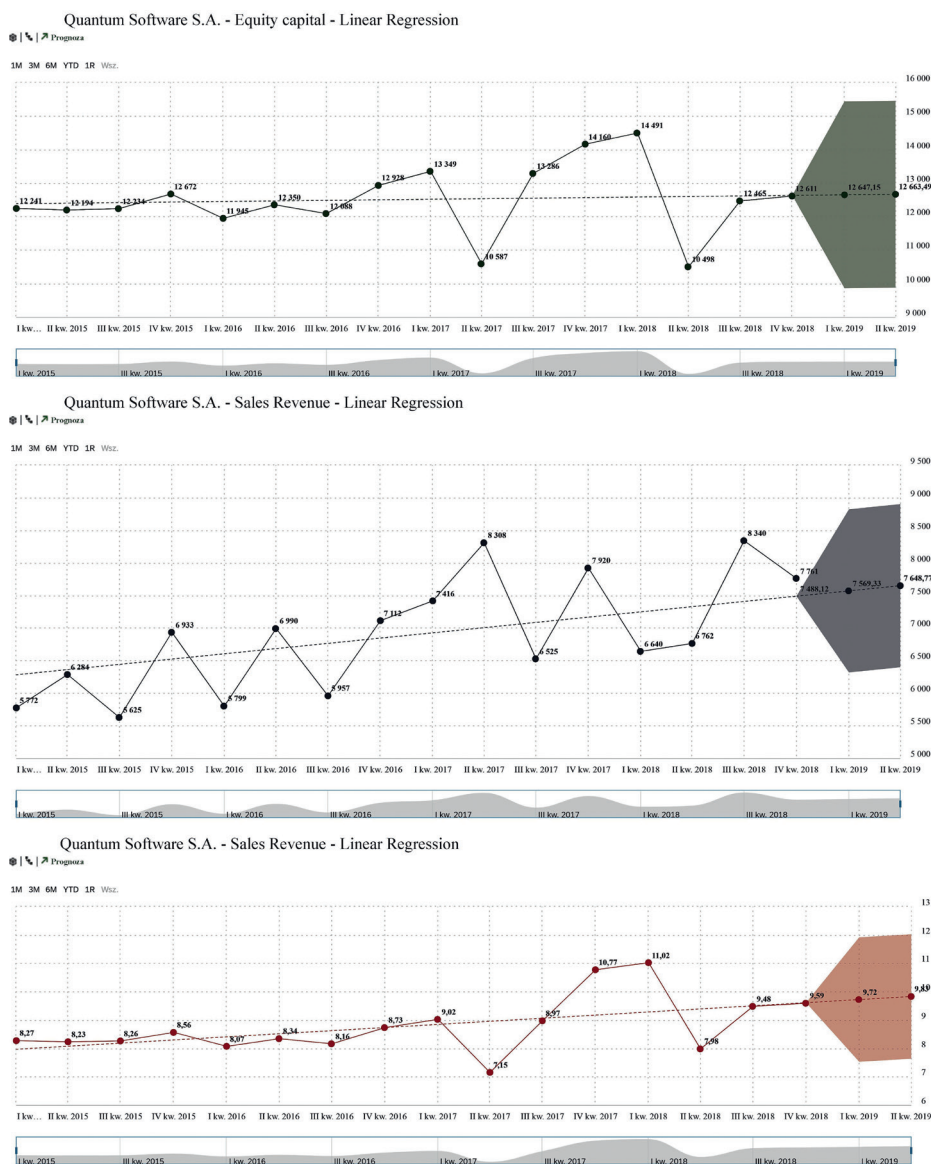


Figure 10. Quantum Software S.A., method: linear regression.

Source: SAC system report.

The charts of equity and book value of 1 share show peaks similar to the charts of equity in Figures 2–5. However, upon careful analysis, you will notice that these charts are similar and have peaks for the same quarters. It is easy to notice the correlation of the book value of 1 share with equity. This is shown in

Figure 11. On the other hand, the sales revenue chart with many peaks suggests seasonality of sales.

The fact that the data is not defective is also evidenced by relatively small differences between the subsequent quarters. In the previous examples, the differences contained in an order of magnitude.

Quantum Software S.A. Comparison of changes in equity capital and book value of 1 share

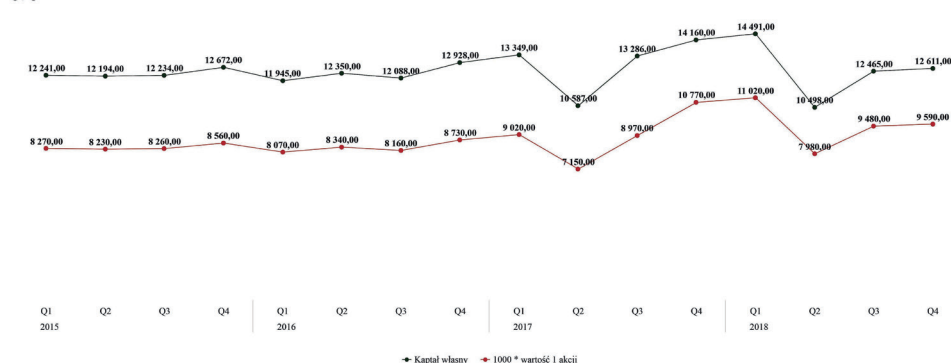


Figure 11. Quantum Software S.A., comparative chart.

Source: SAC system report.

Since the book values of 1 share are about a thousand times lower than the value of equity, they have been scaled in the diagram in Figure 11, multiplied by 1000, otherwise, with the common scale of both graphs, the share values would appear in a straight line. After the scaling operation, the similarity of the graphs for the data under consideration is clearly visible.

To sum up, after analysing all the described cases, it should be stated that the applied CA technique is not a reliable instrument for detecting data defects, but it is a very good "litmus test" for determining a potential threat in terms of ensuring the required reliability.

5. Conclusions and recommendations

Based on the analysis of the subject-matter literature and simulation modelling, taking into account real and simulated business data, it was found that:

1. The quality of the information management system in the company is determined both by ensuring high efficiency at the level of the information product and the process itself (Martyniak).
2. A necessary condition to prevent the appearance of the so-called “Information disease” is the implementation of the concept of 8R (Mangan, Lalwani, Butcher) and 5ZRR (Wiktorowska-Jasik).
3. The most important reasons for the occurrence of data defects include lack of completeness or correctness at the stage of their introduction, lack of updating, field mismatch, conversions between systems or processing errors (Janicki).
4. As a result of the simulation based on the technique “Check Alert”, there is a real benefit in successfully detecting potential data defects.
5. Potential indications of anomalies in the field of data visualization do not automatically mean a data defect and need to be consulted with the data owner.
6. Defect assessment depends largely on the type of a potential defect that requires individual diagnosis, e.g. data values determined by peaks and outside the indicated range or repeated values, etc.
7. In terms of mass handling of business data, a programme of the Business Intelligence class, e.g. the SAC system produced by SAP, is an extremely valuable tool.
8. There are no known intelligent algorithms that could automatically signal data defects in the form of graphical alerts (signalling icons).
9. From the perspective of the efficiency of the information system management quality in the company, it is much better to ensure high standards at the place where the data is generated (data source) and then to detect and correct them (the principle of “zero defects”²⁰ [Crosby]²¹).

To sum up, detecting data defects is not an easy task, especially that there are no effective and automatic methods for detecting data anomalies despite the use of IT systems. However, it should be remembered that incorrect data lead to wrong decisions. Since within the last years there has been a shift in the scope of duties of individual jobs from entering or verifying data towards making decisions,

²⁰ The key principle of TQM “Do it right the first time”. Source: H.H. Steinbeck, *Total Quality Management. Kompleksowe Zarządzanie Jakością*, Wydawnictwo Placet, Warszawa 1998, p. 101.

²¹ E. Skrzypek, *Jakość i efektywność*, UMCS, Lublin 2000, p. 100.

it is important to equip the user/employee with data defect detection techniques through the observed anomalies in their visualization. Despite the introduction of AI systems to decision making, there are still many areas where humans cannot simply be replaced.

References

- Balcerak A., Kwaśnicki W., *Modelowanie symulacyjne systemów społeczno-gospodarczych: różnorodność podejść i problemów*, [in:] A. Balcerak, W. Kwaśnicki (eds.), *Symulacja systemów społeczno-gospodarczych*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2005, pp. 6–7
- Beaverstock M., Greenwood A., Lavery E., Nordgren W., *Applied simulation. Modeling and analysis using FlexSim*, FlexSim Software Products, Orem, Utah, 2011
- Beynon-Davies P., *Database systems*, Macmillan Press Ltd., 1996
- Janicki W., *Jakość bazy danych*, "Automatyka", 9 (2005), item 5; <http://journals.bg.agh.edu.pl> (accessed on July 27, 2019)
- Mangan J., Lalwani Ch., Butcher T., *Global Logistics and Supply Chain Management*, Wiley & Sons, New York – London 2008
- Martyniak Z., *Podstawy diagnozowania informacji jako produktu i procesu*, [in:] Z. Martyniak (ed.), *Zarządzanie informacją i komunikacją. Zagadnienia wybrane w świetle studiów i badań empirycznych*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 2000, pp. 7–100.
- Skrzypek E., *Jakość i efektywność*, UMCS, Lublin 2000
- Steinbeck H.H., *Total Quality Management. Kompleksowe Zarządzanie Jakością*, Wydawnictwo Placet, Warszawa 1998
- Wiktorowska-Jasik A., *Organizacja procesu transportu ładunków transportem zintegrowanym*, [in:] I.N. Semonov (ed.), *Zintegrowane łańcuchy transportowe*, Difin, Warszawa 2008, pp. 42–67

References (online)

- Błaszczuk K., Knosala R., *Problem jakości danych w hurtowniach*, "Prace Naukowe Akademii Ekonomicznej w Katowicach", (2006), <http://www.swo.ae.katowice.pl> (accessed on July 27, 2019)
- Gałęzowski D., Sienkiewicz M., *Zarządzanie jakością danych*, presentation in Warsaw SGH 06/12/2016, <https://studylibpl.com/doc/684763/jako%C5%9B%C4%87-danych> (accessed on July 27, 2019)

Aneta Januszko-Szakiel
ORCID No. 0000-0002-8701-9411
Jagiellonian University

Individual information management. Personal digital archives

Abstract

The article discusses selected issues related to personal digital archives and the management of information resources collected for personal use. Basic concepts were defined: Personal Information Management (PIM) and Personal Digital Archiving, Personal Digital Archiving (PDA). The study presents the basic recommendations for dealing with various types of information, developed by digital curators of archivists, librarians and museum professionals. The discussed elements of the qualitative research conducted using the interview method show a specific example of how large the area of responsibility and work rests on each person who in individual resources gathers documents of their life.

Introduction

The private aspect of gathering information, managing it by individuals in various roles and circumstances of their everyday life, as well as permanent archiving of personal information, became the subject of theoretical considerations and research at the end of the 20th century¹. The studies concerned, among other things, behaviours related to the management of personal collections of information and their permanent protection, identification of groups of users and their competences in the protection of their own digital information collections, the care of institutions obliged the scholars both to conduct teaching and popularize knowledge in this regard. However, there are still no studies showing general regularities in the methods of managing private digital information collections². The personal effectiveness of the activities is, on one hand, the result of the level of awareness and competence of private persons regarding the “handling” of their digital objects, and on the other hand, it is determined by the offer and quality of work of professional digital curators (archivists, librarians, museologists), whose role is connected to being advisers and educators visiting their users in the subject of managing private collections of information in digital form.

The subject of this chapter is a human being and their individual approach towards the management of private digital information resources, collected and at their disposal in order to reuse them, including their long-term storage, and passing them on to future generations. The chapter contains definitions of the following terms: personal information management (PIM) and personal digital archiving (PDA). The recommendations are cited developed in the organizational units of The Library of Congress (LoC) on dealing with private collections of digital resources. The chapter ends with a short report on the author’s qualitative research conducted using the standardized interview technique.

Individual information management

The term Personal Information Management (PIM) is translated and appears in Polish language studies of the subject literally as: “personal information management”, “individual information management”, “own information management”. PIM refers to a variety of information resources that are collected, organized, and maintained

¹ K. Materska, *Zarządzanie informacją i wiedzą*, [in:] *Nauka o informacji*, Babik W. (ed.), Wydawnictwo SBP, Warszawa 2016, pp. 359–385.

² A. Januszko-Szakiel, P. Korycińska, *Osobiste archiwa cyfrowe. Indywidualne praktyki i narzędzia długotrwałego zarządzania zasobami cyfrowymi*, [in:] *Zarządzanie informacją*, W. Babik (ed.), Wydawnictwo SBP, Warszawa 2019, pp. 573–594.

by individuals for later reuse. Individual or personal information is information belonging to an entity (information for the individual and about the entity). The term should be understood as involving both personal data (date of birth, marital status, profession, PESEL number) as well as any other information collected by any persons, encoded in the form of e.g. notes, email messages, web bookmarks, photos, videos, faxes³. The attention of PIM researchers is focused on observing, identifying and characterizing individual human practices while generating, searching, acquiring, classifying, developing, storing, using and distributing information encoded in various forms, in order to use the information in a specific place, time and form, and to achieve the planned goals⁴. The research involves all information resources, both traditional and digital, written by hand, with the help of certain devices, printed on paper, originally digital (“digital born”) and digitized. One of the most important strategies for managing personal information resources is the digitization of traditional (printed, analogue) information resources into digital form. Hence, one of the PIM’s issues is personal digital archives, i.e. individual resources of digital information – collected, organized and permanently protected in order to ensure the possibility of reading and interpreting their content in the distant future⁵.

Personal digital archives

Personal Digital Archiving, Personal Digital Archives (PDA) are terms that define the activities consisting in organizing private digital archives and the permanent

³ W.P. Jones, *Personal Information Management*, “Annual Review of Information Science and Technology”, 41 (2007), 1, pp. 453–504; W.P. Jones, J. Teevan, *Personal Information Management*, University of Washington Press, Seattle 2007; M.W. Lansdale, *The psychology of personal information management*, “Applied Ergonomics”, 19 (1) (1998), pp. 55–66; K. Materska, *Wymiary zarządzania informacją indywidualną*, [in:] B. Sosińska-Kalata, E. Chuchro (eds.), *Spółeczeństwo i sieć informacyjna*, Wydawnictwo SBP, Warszawa 2012, pp. 65–79; R. Sapa, *Metodologia badań indywidualnego zarządzania informacją: wybrane aspekty*, “Zagadnienia Informatyki Naukowej”, 54 (1) (2016), pp. 7–20; M. Świgoń, *Zarządzanie wiedzą i informacją: podstawy teoretyczne, badania w wymiarze indywidualnym*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn 2012, pp. 191–272; J. Tomaszczyk, *Zarządzanie informacją osobistą*, [in:] D. Pietruch-Reizes (ed.), *Zarządzanie informacją w nauce*, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2008, pp. 134–146.

⁴ A. Januszko-Szakiel, *Indywidualne zarządzanie informacją a indywidualna wydajność pracy w przedsiębiorstwach: koncepcja badań*, [in:] P. Cabała, M. Tyrańska (eds.), *Zarządzanie organizacjami w społeczeństwie informacyjnym: innowacje, projekty, procesy*, Instytut Organizacji i Zarządzania w Przemysle ORGMASZ, Warszawa 2017, pp. 160–172.

⁵ A. Januszko-Szakiel, *Archiwistyka cyfrowa: długoterminowa ochrona dziedzictwa nauki i kultury*, Wydawnictwo SBP, Warszawa 2017.

protection of personal content stored in the binary code. PDA is one of the areas of individual information management and digital archiving that focuses on researching and characterizing phenomena concerning the sustainable usefulness of private resources of digital information⁶. PDA issues include the following topics:

- the places and the methods of storing private digital resources;
- the methods of describing and organizing personal digital resources;
- the reasons why people create and protect personal archives of digital resources;
- the competences of the individuals who create digital object collections for their own use and want to manage them as effectively as possible;
- the duties of professional curators of digital resources (librarians, archivists, museologists), who play the role of advisers and educators of people from various backgrounds regarding the care of their personal digital files and the criteria for assessing and selecting private content that should be preserved permanently⁷.

The motivation to handle the issue of organizing and managing personal digital archives stems from the fact that every day, around the world, and in very different situations encountered in private life, people generate, transmit, process and save content that is important to them using digital devices. These contents are “born” in a digital form and the vast majority of them remain in such a digital form, and circulate without printed, analogue surrogates. There is no doubt that the durability of the information recorded in a digital form is much lower than that of that recorded on paper. Reading the content stored on the first types of floppy disks is nowadays an impossible task in the vast majority of cases, and this applies not only to private resources⁸. Therefore, numerous memory institutions, mainly archives

⁶ G. Redwine, *Personal Digital Archiving. DPC Technology Watch Report 15–01 December 2015*, Digital Preservation Coalition. DPC Technology Watch Series, p. 2, <https://www.dpconline.org/docs/technology-watch-reports/1460-twr15-01/file> (accessed on February 20, 2019); M. Wilkowski, *Od osobistej archiwistyki cyfrowej do edukacji medialnej*. „Biuletyn EBIB”, 6 (151) (2014), <http://open.ebib.pl/ojs/index.php/ebib/article/view/274/436> (accessed on February 20, 2019); C.C. Marshall, *Challenges and Opportunities for Personal Digital Archiving*, Chapter in I, Digital, “Chicago Society of American Archivists” (2011), pp. 90–114, <https://www.csd.tamu.edu/~cathycmarshall/I-Digital-Marshall.pdf> (accessed on February 20, 2019).

⁷ A. Januszko-Szakiel, P. Korycińska, *Osobiste archiwa cyfrowe*, pp. 573–594.

⁸ A. Januszko-Szakiel, *Archiwizacja elektronicznych zasobów bibliotecznych. Przegląd stosowanych metod ochrony*, [in:] *Tradycja i nowoczesność w bibliotece naukowej XXI wieku*, A. Januszko-Szakiel (ed.), Oficyna Wydawnicza AFM, Kraków 2012, pp. 131–149; https://repozytorium.ka.edu.pl/bitstream/handle/11315/214/Januszko-Szakiel_Aneta_Archiwizacja_elektronicznych_zasob%c3%b3w_bibliotecznych_2012.pdf?sequence=1&isAllowed=y (accessed on February 20, 2019).

and libraries, also museums, develop and promote catalogues of good practices and sets of recommendations, publish guides, offer training courses, workshops and other forms of individual and group counselling in the field of individual digital heritage management⁹.

Recommendations for dealing with digital resources

One of the first in the world to deal with the issue of personal digital archives were the employees of the Library of Congress of the USA (LoC), explaining that in almost every home there are resources of information that may one day become important and even priceless for a historian, cultural scientist and sociologist. They explain why it is worth it and teach how to transfer the tradition of keeping albums, collecting letters and other types of documents into the digital realm. In addition to raising awareness about the need to create personal digital archives, LoC librarians provide their users with certain guidelines supporting them in the tasks of archiving various types of private digital resources¹⁰.

Archiving of digital photos

Digital photos are a particularly important type of personal digital resources because of the emotions that come with such photos, but cannot be recreated or replaced if lost. LoC librarians recommend their users to do the following:

- search for all digital photos in personal cameras, computers and removable media such as memory cards, also adding the photos stored on the network;
- select images that your users think are worth keeping – if there are multiple versions of an important photo, save one at the highest quality;
- give photos descriptive names, label them with the names of people and descriptions of events;

⁹ *Personal Digital Archiving Day Kit*, Library of Congress, <http://digitalpreservation.gov/personalarchiving/padKit/index.html> (accessed on February 20, 2019); *Zostań rodzinnym archiwistą*, Naczelna Dyrekcja Archiwów Państwowych <https://www.archiwa.gov.pl/pl/aktualnosci/5030-warsztaty-%E2%80%9Ezosta%C5%84-rodzinnym-archiwist%C4%85%E2%80%9D> (accessed on February 20, 2019).

¹⁰ *Preserving Your Digital Memories The National Digital Information Infrastructure and Preservation Program. A collaborative Initiative of The Library of Congress*, http://www.digitalpreservation.gov/personalarchiving/documents/PA_All_brochure.pdf (accessed on February 20, 2019).

- group photos thematically and create an appropriate folder structure on your computer for them, give the folders names that indicate their content;
- make at least two copies of a set of selected photos;
- store each copy in a different physical location and preferably on a variety of media types (DVD, CD, portable hard drives, USB drives, or network drives); if a disaster affects one of the carriers where the photos are stored, the copy stored elsewhere should be safe;
- check their photos no less than once a year in order to make sure that they are readable; for copies made on physical media, you should refresh the copies on new media every five years (or more frequently if necessary) to avoid data loss.

Archiving digital audio recordings

Some of the audio files of music, lectures and other recordings that people own are of special personal importance to them, encouraging the owners to preserve them in long term. It is necessary to make sure that the audio files selected to be stored are in an open format. This will give you the best chance of seamlessly recreating the data in the future. The librarians' recommendations included the following procedures:

- searching for all digital audio files on personal computers, voice recorders, telephones and removable media such as memory cards or optical carriers;
- deciding which audio recordings are worth storing – if there are multiple versions of an important recording, save one in the highest quality, in an open format;
- giving descriptive names to files and grouping them thematically;
- labelling files with the information about the recording;
- creating an appropriate folder structure on the computer to place selected recordings in it and giving folders names that describe their content;
- making no less than two copies of a set of selected audio recordings;
- storing each copy in a different physical location and preferably on different types of carriers (DVD, CD, portable hard drives, USB drives or network drives);
- checking your personal audio files not less than once a year in order to make sure they are readable; for copies stored on physical carriers, it is recommended to refresh the copies on new media every five years (or more often if needed) to avoid irretrievable data loss.

Archiving personal digital video records

Using a video camera or other digital image capture device means that the tradition of making personal and home movies is being continued. You may want to keep some of these movies for a long time. For digital video, the technical quality of the files is an important factor. It is a good idea to save your private videos at the highest quality available (for example, videos posted online are often “grainy” and provide much less detail than the original version) with relevant descriptive information (metadata) about such videos. The tips given by professionals for archiving video recordings as another type of personal digital object are similar to the guidelines mentioned above; the professionals point out to the need to do the following:

- search for all digital films in private cameras, photo cameras, telephones, removable media, as well as attaching films stored in the Internet space;
- select movies that in users’ opinion are worth keeping and save one file from among numerous versions in the highest quality;
- give files with films descriptive names, mark them with the names of people and descriptions of events;
- group the videos thematically and create an appropriate folder structure for them on the computer, give the folders names that describe their content;
- make at least two copies of a set of selected video recordings;
- store each copy in a different physical carrier and on different types of media;
- check the possibility of playing private movies not less than once a year to make sure they can be read; for copies made on physical media, refreshing of copies on a new medium should be performed every five years.

Archiving personal digital correspondence

Just like hard-copy letters, emails record important events, transactions, and relationships. Archiving email messages requires separating them from the managing email client. This is because email clients are not designed to store information for a long time; they are only used for efficient and ongoing management. Therefore, librarians urge the users to do the following:

- identify all personal email accounts and include all folders or other dedicated message groups in each one; including all “archived” resources;
- decide which of the messages are worth keeping – if the messages contain attachments, do not forget about the attachments;

- export selected messages from the email client to separate files; if possible, save messages in an open format, preserving also the message metadata, including the message “header” (subject, date, time);
- give email files descriptive names;
- group the messages thematically and create an appropriate folder structure for them on the computer, give the folders names that describe their content;
- make at least two copies of the emails designated for storage;
- store each copy in a different physical carrier and on different types of carriers;
- check the possibility of reading stored messages at least once a year; for copies made on physical carriers, it is a good idea to refresh them on new media every five years (or more often if needed).

Archiving documents recording personal life

Probably every private digital collection contains different versions of resumes, school papers, financial documents, medical records, presentation slides, or digital copies of original paper documents that may be of permanent value. With information of that kind, it is important to decide which documents to store, and to think about different versions of the documents, including draft and back-up copies. Drafts can contain important details that do not appear in final versions – sometimes it is useful to keep both the draft and the final document. When starting the process of archiving this type of resources, carefully search through all digital documents on computers, removable media, networks, and follow the recommendations mentioned earlier for other types of digital objects.

Archiving personal websites, blogs and social media

Running a blog, website, or social media accounts requires considering whether and to what extent it is worth saving the information published there for the future. For this type of information, the archiving process should start by identifying all the places where the information is published and considering all of them, both current and older and archived. It is recommended that you export the content selected for permanent storage in order to separate files or a set of interrelated files, and also to keep metadata with the information about the WWW content, such as the name of the website or the date of creating. The remaining archiving procedures are typical and do not differ significantly from those already mentioned for other types of information.

Very similar recommendations have been developed within the structures of Digital Preservation Coalition (DPA)¹¹. The DPA lists several sets of practices (basic and extended) of individual information management over a long time, highlighting that the system may be expanded and be the basis for freely constructed and even most sophisticated individual strategies for managing private resources of information. Implementing only a basic list of recommendations into practice is supposed to protect usefulness, i.e. accessibility and readability of the content of individually collected digital resources¹².

Interview – research report

The research topic presented in this section originates from the qualitative research conducted by the chapter's author within the years 2018–2019. The research tool was an interview questionnaire consisting of seven questions. The aim of the study was to determine whether, how and with what effects, people of different ages, with different levels of education and with different level of digital competences, cope with preserving the availability and the usefulness of the collected private resources of digital information.

One of the participants of the study was a 37-year-old man with a university degree in economics and supplementary education on information sciences. Professionally, he worked as an editor of a digital repository, he was also involved in project management.

During the interview, it was established that the interviewee's private digital resources include photos, videos, music, text and graphic files, digital notes, and email messages. Most of the objects (private photos and videos, text and image files) are stored in the cloud (Google Drive and Microsoft OneDrive), some of the most important ones on an external hard drive (which serves as an archive and backup copy). The respondent admitted that he reviewed private digital resources "mostly ad hoc and selectively", mainly "if he wanted to use a given resource". He is no stranger to loss of access, i.e. being able to read his own digital objects. This "happened when I was still using DVDs to archive my resources. A few DVDs got scratched, making the resources unreadable and therefore unreachable". In spite of having such an incident, the interviewer openly admitted: "I am not reviewing the external drive's performance regularly".

¹¹ G. Redwine, *Personal Digital Archiving*, p. 2.

¹² A. Januszko-Szakiel, P. Korycińska, *Osobiste archiwa cyfrowe*, pp. 573–594.

The further part of the interview indicated that the interlocutor is aware of the need for a well-thought-out and planned management of private resources of information and is characterized by a level of digital competence adequate for these tasks. Even so, there is an evident gap in his personal digital information management that consists in not reading certain digital information objects for too long. The interviewee declared that he knew the term “individual information management”, while he was not familiar with “personal digital archives”. When asked about his individual strategy, meaning the methods and tools used, of managing private digital resources, he replied: “I listen to music using SaaS software (Spotify). For digital information, I use the Evernote note management program. I manage my email correspondence through the Windows Mail client configured using the automatic synchronization protocol (IMAP) on each of the four «connected» email accounts”. The interlocutor chose the way of dealing with digital information on the basis of self-study and gathered knowledge (mainly from the Internet sources). “Training and education (business info-brokering), professional experience (6 years as an editor of a digital repository)” proved to be helpful. The respondent did not know the services offered by science and culture institutions (archives, library, museum) as regards courses, workshops, other forms of education and consultancy on the management of own/private information collections. He came across “the offer of an institution, which contained certain elements or concerned specific types of digital resources (e.g. files in the form of scientific publications), but did not cover the entire issue of managing one’s own collections”.

The interview presented above, although only published in fragments, clearly indicates that the problem of digital content loss applies to every person who generates and stores it for their own use. Managing private resources of digital information is a task that emerges in people’s daily lives. Regardless of age, gender, education level, profession, people live in an environment of digital content and devices for generating and storing such content. They register and collect information – sometimes the information is of particular importance, in other cases – hardly significant. Not everyone has the knowledge and skills, self-discipline and self-motivation to effectively “care for” their collections of personal information. You need the support of the professionals who are involved in the development and teaching of good practice in this area. Above all, it is important to mention an individual responsibility of every person for the storage of private information, which may, both now and in the future, be not only of value for those involved, but also be a valuable supplement to institutional collections, put to use by the general public, contributing to the collection digital resources of the region, country and the world.

References

- Januszko-Szakiel A., *Archiwistyka cyfrowa: długoterminowa ochrona dziedzictwa nauki i kultury*, Wydawnictwo SBP, Warszawa 2017
- Januszko-Szakiel A., *Indywidualne zarządzanie informacją a indywidualna wydajność pracy w przedsiębiorstwach: koncepcja badań*, [in:] Cabała P., Tyrańska M. (eds.), *Zarządzanie organizacjami w społeczeństwie informacyjnym: innowacje, projekty, procesy*, Instytut Organizacji i Zarządzania w Przemśle ORGMASZ, Warszawa 2017, pp. 160–172
- Januszko-Szakiel A., *Archiwizacja elektronicznych zasobów bibliotecznych. Przegląd stosowanych metod ochrony*, [in:] *Tradycja i nowoczesność w bibliotece naukowej XXI wieku*, Januszko-Szakiel A. (ed.), Oficyna Wydawnicza AFM, Kraków 2012; https://repozytorium.ka.edu.pl/bitstream/handle/11315/214/Januszko-Szakiel_Aneta_Archiwizacja_elektronicznych_zasob%c3%b3w_bibliotecznych_2012.pdf?sequence=1&isAllowed=y (accessed on February 20, 2019)
- Januszko-Szakiel A., Korycińska P., *Osobiste archiwa cyfrowe. Indywidualne praktyki i narzędzia długotrwałego zarządzania zasobami cyfrowymi*, [in:] *Zarządzanie informacją*, Babik W. (ed.), Wydawnictwo SBP, Warszawa 2019
- Jones W.P., *Personal Information Management*, “Annual Review of Information Science and Technology”, 41 (2008), 1, pp. 453–504
- Jones W.P., Teevan J., *Personal Information Management*, University of Washington Press, Seattle 2008
- Lansdale M.W., *The psychology of personal information management*, “Applied Ergonomics”, 19 (1) (1998), pp. 55–66
- Materska K., *Wymiary zarządzania informacją indywidualną*, [in:] *Spółeczeństwo i sieć informacyjna*, B. Sosińska-Kalata, E. Chuchro (eds.), Wydawnictwo SBP, Warszawa 2012, pp. 65–79
- Materska K., *Zarządzanie informacją i wiedzą*, [in:] *Nauka o informacji*, Babik W. (ed.), Wydawnictwo SBP, Warszawa 2016, pp. 359–385
- Sapa R., *Metodologia badań indywidualnego zarządzania informacją: wybrane aspekty*, “Zagadnienia Informatyki Naukowej”, 54 (1) (2016), pp. 7–20
- Świgoń M., *Zarządzanie wiedzą i informacją: podstawy teoretyczne, badania w wymiarze indywidualnym*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn 2012, pp. 191–272
- Tomaszczyk J., *Zarządzanie informacją osobistą*, [in:] D. Pietruch-Reizes (ed.), *Zarządzanie informacją w nauce*, Wydawnictwo Uniwersytetu Śląskiego, Katowice 2008, pp. 134–146

References (online)

- Marshall C.C., *Challenges and Opportunities for Personal Digital Archiving*, Chapter in I, Digital, "Chicago Society of American Archivists", (2011), pp. 90–114, <https://www.csdl.tamu.edu/~cathycmarshall/I-Digital-Marshall.pdf> (accessed on February 20, 2019)
- Preserving Your Digital Memories The National Digital Information Infrastructure and Preservation Program. A Collaborative Initiative of The Library of Congress*, http://www.digitalpreservation.gov/personalarchiving/documents/PA_All_brochure.pdf (accessed on February 20, 2019)
- Redwine G., *Personal Digital Archiving. DPC Technology Watch Report 15–01 December 2015*, Digital Preservation Coalition. DPC Technology Watch Series, p. 2; <https://www.dpconline.org/docs/technology-watch-reports/1460-twr15-01/file> (accessed on February 20, 2019)
- Wilkowski M., *Od osobistej archiwistyki cyfrowej do edukacji medialnej*, "Biuletyn EBIB", 6 (151) (2014), <http://open.ebib.pl/ojs/index.php/ebib/article/view/274/436> (accessed on February 20, 2019)

Mariusz Grzyb
ORCID No. 0000-0001-8439-9650
Andrzej Frycz Modrzewski Krakow University

Dorota Kowalik
ORCID No. 0000-0002-9242-2002
Andrzej Frycz Modrzewski Krakow University

Information security management – audit of the IT system

Abstract

In the article we indicated how internal audit is important in any organization, how it can support the organization in achieving goals. The internal audit can reduce the risks occurring in the organization, while IT audit is an indispensable tool supporting the processes of maintaining business continuity to guarantee the organization's resilience to unforeseen events.

1. Introduction

In order for an organization to survive, it must successfully achieve its basic goals and objectives through continuous development. Considering the constant changes in economic, legal, market, etc., an organization, while conducting its business, is forced to react quickly to such changes, which require appropriate management. Efficient and effective management of an organization is a set of activities that involve planning, making decisions, organizing, managing people and controlling – organization management consists of numerous processes.

It is widely recognized that the first classification of the management functions was formulated by Henri Fayol in the book from 1909. *L'exposee des principes generaux d'administration*. He distinguished five basic functions, which are as follows:

- planning (closely related to forecasting);
- organizing (i.e. all administrative activities);
- coordinating (resources and activities);
- leadership (ordering, giving orders, leading to the implementation of the plan);
- controlling (implementation of the plan).

Among the elements of management Peter Ferdinand Drucker – an expert in management, lists an extensive system of indicators that allow us to constantly and comprehensively (taking into account the criteria important for the entity) monitor (test), evaluate and correct (improve) the quality, rationality, efficiency and effectiveness (efficiency) of operations¹³.

The results of the work of only two scientists have been quoted as an example, but it is noteworthy that many scientists mention controlling or monitoring among the basic management functions. It can be noted that organizations have developed many types of control, including internal audit.

A significant change in the role played by audit in an organization occurred at the beginning of the 20th century, with the emergence of international corporations which with time wanted to reduce their dependence on statutory auditors verifying only the correctness of accounting records and the compliance of accounting procedures with the regulations, and wanted to verify the effectiveness of accounting controls by employing their own internal auditors, who reported to the management of the organization. Since then, the evaluation of the role of internal audit has commenced. Currently, internal audit is regarded as an essential element

¹³ S. Duchniewicz, *Metody organizacji i zarządzania*, PTM, Warszawa 2005, pp. 15–22.

of the assessment of the risk of organization management, it is a management tool used to assist the organization's managers to rest assured that

- the goals set for the organization are being achieved;
- procedures in the organization originating from applicable legal regulations are properly implemented and followed;
- mechanisms and procedures are adequate and effective for the proper functioning of the organization.

Internal audit supports an organization in achieving its goals by systematically and consistently improving the effectiveness of the organization's management¹⁴.

Currently, the basic document that defines and regulates internal audit are "International Standards for the Professional Practice of Internal Auditing compiled and published by The Institute of Internal Auditors (IIA)"¹⁵. According to the currently applicable definition, internal audit "is an independent and objective activity aimed at adding value and improving the organization's operational activity. It consists in a systematic and orderly evaluation of the processes: risk management, control and governance, and contributes to the improvement of their functioning. It helps the organization in achieving its goals by providing assurance about the effectiveness of said these processes as well as through consulting"¹⁶.

In order to standardize the audit work, the Institute of Internal Audit (IIA) has developed standards for professional internal audit practice, including:

- standards of attributes defining the organization's characteristics and persons carrying out the audit;
- operating standards, outlining the types of the internal audit activities and defining the qualitative criteria of the assessment;
- standards of implementation related to specific types of audit engagements (for example, compliance audits).

On the other hand, according to Polish legal regulations, internal audit is governed primarily by the Public Finance Act, which regulates the general scope of internal audit in section VI, entitled "Internal audit and coordination of internal audit in state financed units", according to which, "internal audit is an independent and objective activity, the purpose of which is to support (...) the head of the entity

¹⁴ K. Czerwiński, *Książka procedur audytu wewnętrznego*, CISA, Warszawa 2002, unpublished materials, p. 7.

¹⁵ International Standards for the Professional Practice of Internal Auditing, The Institute of Internal Auditors, www.ii.org.pl.

¹⁶ Definition of internal audit, Code of Ethics and International Standards of Professional Practice of Internal Auditing according to Institute of Internal Auditors, p. 3.

in achieving goals and tasks through systematic assessment of management control and advisory activities. This assessment relates in particular to the adequacy, effectiveness and efficiency of management control in (...) the organization”¹⁷.

The provisions of the applicable state legal regulations are more and more harmonized with the provisions of international regulations regarding the approach towards internal audit issues. There is a strong emphasis on supporting the head of the organization in the realization of its goals and tasks through examination and assessment of management control (assurance activity) and consulting (advisory activity). What we can notice here is a holistic aspect consisting in control and management (covering the entire economy within the unit).

The question is why we have chosen the subject of IT audit for our diploma thesis. In the era of computers and networks, IT systems depend on virtually every aspect of business in any organization, regardless of the size of the business or industry. All business and administrative processes are held within the IT systems. Data confidentiality and integrity, access control systems, and video monitoring systems all depend on IT infrastructure. Therefore, it is difficult to identify an area that is more sensitive and that influences business security. So, an IT audit is an indispensable tool supporting the processes of maintaining business continuity to guarantee the organization’s resilience to unforeseen events.

2. Information security

Last years have seen a very dynamic development of information technologies, and hence, growing threats in the area of information security. Loss of confidentiality, access limitation or breach of integrity are just some of the threats to information security. Threats may be brought about by natural factors (electrical overvoltage, flooding, fire), accidental (ordinary failures or errors, e.g. in software) and, which is more and more often, human deliberate actions (cyber attacks, embezzlement, burglary and data theft).

Information security is extremely important in the context of business activities. Numerous organizations focus on ensuring information security by implementing security procedures and policies based on accepted standards. Nowadays, the (ISO/IEC) standards are considered the foundation of information security solutions. Information security has also penetrated legal documents. As many as four standards have been included in the Notice of the Prime Minister of November 9, 2017 on the publication of the consolidated text of the Regulation of the Council

¹⁷ The Act of 27 August, 2009 on Public Finances (Polish Journal of Laws of 2019, item 2020).

of Ministers on the National Interoperability Framework¹⁸; minimum requirements for public registers and exchange of information in electronic form, and minimum requirements for ICT systems. The requirements set out in this notice are deemed satisfied if the system for information security management has been developed according to the Polish Standard PN-ISO/IEC 27001, and implementing safeguards, risk management and auditing are carried out according to the subsequent ISO/IEC 27000 standards.

The PN/ISO IEC 27001 standard contains over one hundred and twenty principles regarding information security grouped into fourteen areas:

- Information security policy;
- Organization of Information Security;
- Security of human resources;
- Asset management;
- Access control;
- Cryptography;
- Physical and Environmental Security Policy;
- Safe exploitation;
- Communication security;
- System acquisition, development and maintenance;
- Relationships with suppliers;
- Management of information security incidents;
- Information security aspects of business continuity management;
- Compliance.

The methodology we have developed took into account the assessment of the compliance of the university system with the requirements of ISO 27001 as the basis for the audit. We chose this standard due to the fact that ISO 27001 helps to protect information in several aspects:

- confidentiality – only authorized persons may access information;
- integrity – ensures the accuracy and completeness of information and processing methods;
- availability – authorized users can access information and related assets when required.

¹⁸ Regulation of the Council of Ministers of 12 April 2012 on the National Interoperability Framework, minimum requirements for public registers and electronic information exchange, and minimum requirements for ICT systems (Polish Journal of Laws: Dz.U. of 2012, item 526).

3. Internal audit

Before proceeding to the identification of the most important elements of our diploma thesis, we will briefly present individual elements of the internal audit.

In the Polish subject-matter literature, the most common are the following three types of internal audits:

- financial audit;
- operational audit;
- IT audit.

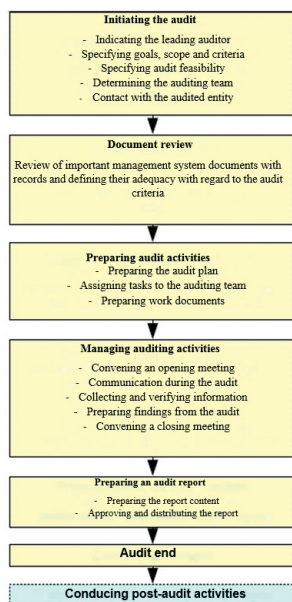
Taking into account that in this diploma paper the audit has covered the IT aspect, we will focus here only on this type of audit, adding that an IT audit can be regarded as an independent undertaking or as an element of a financial or operational audit. The said audit covers all IT systems and related resources (i.e. buildings, energy systems, cable systems, air conditioning systems that are necessary for the operation of IT systems, all system platforms, database systems, application software, computer hardware, problems related to management of software licenses, problems related to risk management processes and the process of ensuring system security).

It should be highlighted that the IT audit is used to examine and to assess the correctness and the effectiveness of internal control system for various processes applied within the organization. The purpose of conducting this type of audit is to improve the functioning of the organization by increasing the efficiency of decision-making and of the business processes. IT audit is a process that is designed to determine whether IT systems and the people responsible for them properly protect data, thus guarantee the security of business processes and effectively help to achieve the goals of the organization while ensuring resistance to undesirable events.

IT audit work, just like any other type of audit, should meet certain standards and proceed in a systematic, consistent and orderly manner. IT audit can be divided into four main phases: planning, assessment of control mechanisms, testing and reporting.

3.1 Audit activities

Several key activities are distinguished in the audit, and they are presented in the diagram below:



Scheme 1. Audit activities.

Source: elaboration after https://mfiles.pl/pl/images/c/c0/Audyty_wewnetrzny.png (own translation).

3.1.1 Initiating the Audit. Review of documents. Preparation of audit activities

In order to maximize the use of the resources with minimal disruption to work, proper planning of the audit process is an essential thing. The planning phase is very important so that the audit can run smoothly. Before starting the actual audit work, the auditor focuses on analysing the purpose of the audit, as well as the assumed criteria, and the scope of the audit, which allows for defining the boundaries within which the audit is carried out.

The chief audit executive is responsible for creating an objective audit plan by relying on the results of the risk analysis in determining the audit plan. An audit using its own professional and objective assessment takes into account the performance of the organization's existing risk management system, including information on risk appetite levels set by management for different activities or parts of the organization. If such a system does not exist, the chief audit executive performs a risk assessment after considering information received from senior management and the board. The chief audit executive must review and adapt the plan as necessary to fit organizational changes, risks, operations, programs, systems, and controls.

When planning the task, the auditor takes into account the risks relevant to the activity and the measures taken by management to reduce the risk to an acceptable level. In developing the plan for internal auditing works and determining the priorities necessary to allocate resources, the internal auditor performs risk assessment. Such assessment is used for analysing the audited entities and selecting the most risky areas for the internal audit work plan.

The factors that the internal auditor takes into account in developing the audit plan include the following:

- inherent risks – have they been identified and assessed?
- residual risks – have they been identified and assessed?
- risk mitigation controls, accident plans and monitoring activities – are they related to specific incidents and/or risks?
- risk registers – are they kept systematically, are they complete and accurate?
- documentation – are the risks and activities documented?

An internal audit plan typically focuses on:

- unacceptable current risk when action is required on the part of the management;
- areas where key controls and mitigating factors are minimal and therefore senior management needs to investigate them promptly;
- the systems of control that the organization most relies on;
- the areas with a very large difference between inherent and residual risk;
- the areas where the inherent risk is very high.

When planning a certain task, the internal auditor identifies and assesses the risks concerning a given area. This stadium is also referred to general or initial overview. Planning an audit is primarily information gathering, but it is only the first stage in obtaining information that will be obtained during auditing activities. Information collected at the planning stage can be divided into several categories:

- legal environment – all legal acts, guidelines of external bodies, contracts and agreements relevant to the audited business;
- organizational system – organizational structures on which the operation of the unit is based. The competences and the scope of duties assigned to specific persons holding given positions, terms and conditions of delegating powers, the system of subordination and superiority between the functions, the number of employees and the organization's inventory register;
- financial information – financial plans and reports regarding implementation, income and expenditure plans, records and estimates of operating costs, taking into account past data, which the auditor will use as a benchmark;

- policy, procedures and instructions – they characterize the methods of carrying out the activity, thus they are the source of comparative criteria. This element may itself be the subject of an audit for its impact on savings, efficiency and effectiveness;
- other information – reports from previous audits, inspections, reviews (internal sources) or publicly available studies on a given problem, area, such as scientific publications, press articles and any other information provided by managers.

3.1.2 Carrying out audit activities. Preparation of the audit report

3.1.2.1 Opening conference

An extremely important issue from which the actual audit starts is the opening meeting. During the opening meeting, apart from confirming the scope of the audit and presenting the methodology of its conduct, organizational and technical issues related to the course of the audit are also discussed.

3.1.2.2 Auditing works

Auditing works are carried out using the sampling method, based on the prepared checklists. This means that the entire area is not examined in detail, but just its selected elements. Samples should be corroborated with evidence as they will be used at the reporting stage to formulate the detected non-conformities.

In order to complete the information, the auditor conducts interviews with selected users of the information system. Following the checklist, it systematically collects information and checks whether the necessary security has been identified and properly implemented. On-site inspections are an additional source of information about the system for information security management.

3.1.2.3 Reporting

The results of the audit are presented in the form of an audit report. The recipient of the report is the management of the audited organization. The non-conformities formulated in the report should be corroborated by evidence and should refer to the points of the standard.

The audit report should include the following in particular:

- the details of the auditee;
- the purpose and the scope (areas) of the audit;
- the standard used;
- the description of audit methods;
- the names and surnames of the auditors;
- the findings (samples) and the evidence of these findings;
- checklists;
- indicated non-conformities along, showing the points within the standard where these non-conformities occur;
- recommendations for corrective actions.

The auditee has the right to make reservations and comments to the report. In this case, the lead auditor has to address the comments and the reservations supplied. The report is a confidential document that must also be physically secure.

3.1.2.4 Corrective actions and closure of the audit

The purpose of preparing a summary report is to bring all the facts together so that they can be presented to the auditee at the closing meeting. The summary report should contain a clear interpretation of the facts in order to present the auditee with audit conclusions and to make sure that the auditee understands them.

3.1.2.5 Closing meeting

The closing meeting is a formal meeting that proceeds according to a specific plan. The lead auditor is responsible for discussing the audit work and presenting the summary report.

The agenda of the closing meeting consists of the following elements:
recalling the purpose and scope of the audit;

- discussing the method of reporting – confirmation that the copies of non-conformity reports signed by both parties will be provided after the meeting, and a complete report will be sent within the agreed deadline;
- non-conformity report – presents the non-conformities identified by the auditor
- summaries of audit works – presenting conclusions and recommendations based on the entire audit, described in the summary report;
- explanations – the auditee has the right to receive explanations regarding all aspects of the audit, asks auditors questions and ask for clarification of unclear issues.

3.1.2.6 Post-audit activities

Post-audit activities are mainly an implementation of corrective actions for detected non-conformities during the internal audit.

4. The summary of the most important elements of the diploma paper

In our opinion, a very important element of our diploma paper was the process of identifying and estimating risk due to the fact that it is an inseparable process related to the functioning of internal audit. The process of identification and estimation of risk is part and parcel of the auditor's work at every stage of his work, commencing with strategic planning, through the preparation of annual audit plans, and ending with the implementation of individual audit tasks. Reducing risk in organization's operations requires creating a list of all processes taking place in it and conducting a comprehensive risk analysis for such processes. The risk analysis is a tool for rationalizing the selection of those areas of the organization's operation that generate the greatest risk and require audit in the first place. It should be emphasized that there is no universal set of risky areas, and the identification of risk areas depends on the auditor's professional judgment.

Another very important element of our diploma paper is preparing internal control assessment forms on the basis of which the audit has been conducted.

4.1 Risk

“The principles and the procedure of IT system risk assessment were prepared for the purposes of our study in the light of information security” – hereinafter referred to as “the rules”.

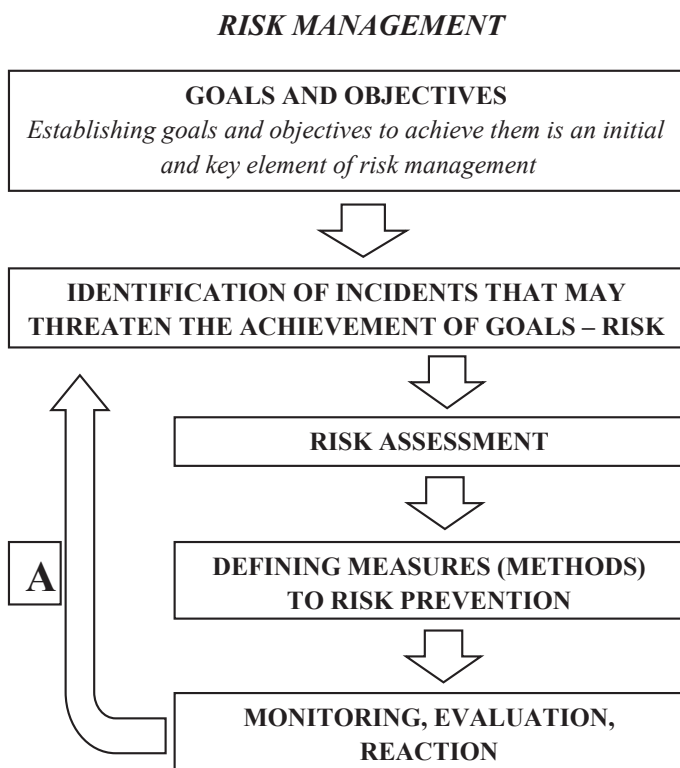
The purpose of drafting the Principles was to establish methodology for risk assessment related to university’s IT-related activities in the light of information security. The Principles defined the manner of conducting and documenting the risk assessment process.

4.1.1 Risk management

Risk management is an activity that increases the level of security in the IT area at a university. The activities undertaken as part of these activities and the technical and the applied organizational measures depend on the environment in which the data is processed.

Risk management of the university’s IT system involves activities consisting in:

- identifying processes;
- identifying the vulnerabilities and the already existing security;
- risk assessment;
- risk handling;
- accepting risk;
- risk monitoring;
- risk communication.



Scheme 2. Risk management.

Source: elaboration after https://www.edurada.pl/assets/_resampled/resizedimage635523-Schemat-zarzadzania-ryzykiem.jpg (own translation).

As a result of the risk estimation process, adequate threats and the probability of their occurrence, as well as the technical and organizational measures have been determined.

4.1.2 Risk estimation

Risk estimation was aimed at determining what might happen (when, where, how and why) and how severe losses may arise therefrom. The risk estimation consisted of:

- identifying threats;
- risk analysis;
- risk estimation.

As part of the risk identification, the following were specified:

- context;
- asset identification;
- hazard identification for the assets;
- identification of the existing security measures;
- identification of vulnerabilities;
- identification of consequences – effects, i.e. loss of control over own data, data destruction, data leakage, including personal data, identity theft or falsification, financial loss, unauthorized reversal of pseudonymization, breach of confidentiality of data protected by professional secrecy and any other significant economic damage or social.

4.1.3 Identification

Having identified assets, threats and applied security measures, an identification of the susceptibility to the occurrence of specific threats has been carried out. Importantly, the mere existence of vulnerability does not yet cause loss. The occurrence of loss is possible only after the threat that will exploit a given vulnerability has materialized. Vulnerability analysis concerns basic assets – processed data and the procedures used for processing – together with auxiliary processes – hardware, software, computer network, employees, headquarters, organization.

4.1.4 Analysis

The three most important elements of a risk analysis are as follows:

- determining the consequences, with a particular focus on the possibility of data breaches, including data, including personal data;
- estimating the likelihood of an incident;
- estimating the level of risk.

The basic type of risk response is action or transfer of risk. Transfer is the transfer of risk to a third party. The action may include, in particular, the establishment of new or intensification of the existing control mechanisms, as well as actions of a different nature (e.g. employee training, introducing organizational changes, requesting additional funds, introducing additional information requirements, taking or intensifying control activities, etc.).

For the purposes of the risk analysis, the following criteria for assessing its individual elements were adopted in the diploma paper:

The estimation of the effects of the occurrence of risk has been defined on a three-point scale, where 3 means a very big effect and 1 – a very low one. These criteria have been presented in Table 1.

Table 1. Estimation of the effects of the occurrence of risk.

Estimating the effects of the occurrence of risk			
Effect	Level	Financial value/loss	Impact on the IT system
Critical	3	PLN 100 – 500 thousand	Affects
High	2	PLN 50 – 100 thousand	Partly affects
Low	1	PLN 5 – 50 thousand	No impact

Source: own elaboration.

The estimation of the probability of risk occurrence was specified on a three-point scale, where 3 means a very high probability of risk occurrence and 1 – a low probability. These criteria have been presented in Table 1.

Table 2. Estimation of the probability of risk occurrence.

Estimating the likelihood of risk occurrence		
Probability	Level	Description of the probability
Probable	3	Occurs no less than once a month
Possible	2	Occurs no less than every 3 months (quarter)
Unlikely	1	Occurs no less than once every 6 months

Source: own elaboration.

The risk significance level will be calculated according to the formula: Risk (R) = effect (S) x probability (P). For the purposes of the analysis, it was assumed that a result of 1–2 means a low level, 3–4 a high level, 6–9 a critical level. The scaling of the risk level is presented in Table 3.

Table 3. Effect.

PROBABILITY	EFFECT				
	S x P = R		Low	Medium	High
			1	2	3
	Probable	3	3 Ś	6 W	9 W
	Possible	2	2 N	4 Ś	6 W
	Unlikely	1	1 N	2 N	3 Ś

Source: own elaboration.

It is assumed that the level of risk classified as high or above is not acceptable. The risk acceptance criteria are presented in Table 4.

Table 4. Risk acceptance criteria and possible actions.

Risk acceptance criteria and possible actions		
Risk level	Acceptable YES/NO	Description of actions taken
Low (N)	YES	Action can be postponed but requires periodic monitoring
Medium (M)	NO	Action required
High (H)	NO	It requires immediate action

Source: own elaboration.

4.1.5 Result

As a result of the risk estimation of the IT area within the university, and taking into account twelve areas of information security, the audit study covered IT areas that are most susceptible to risk:

- security policy;
- physical and environmental security;
- access control.

The results of the initial review have been used to develop the audit task program, i.e. the basic document where the auditor presents the results of the initial review and specifies the planned auditing works.

4.2 Internal Control Assessment Forms

In this diploma paper, we made an assumption that the above-mentioned IT areas would be audited basing on the prepared internal control forms, and taking into account five areas and the accompanying sub-areas of management control, separated in accordance with the Communication No. 23 on management control standards¹⁹, i.e.:

- internal environment;
- goals and risk management;
- control mechanisms;
- information and communication;
- monitoring and evaluation.

The standards indicated above, on which we relied on the preparation of internal control assessment forms for our diploma thesis, although they define the basic requirements relating to management control in the public finance sector and relate to management control in the public finance sector, can be used in auditing in any organization as a coherent and uniform model of management control.

As an example, this article, due to its extensive structure, contains a fragment of the internal control form of the internal environment as regards information security within the organization, on condition that such a form has been prepared separately for the areas indicated as most exposed to risk, i.e.:

- security policy;
- physical and environmental security;
- access control.

¹⁹ Announcement 23 of the Minister of Finance of December 16, 2009 on management control standards for the public finance sector.

Table 5. Excerpt from the internal control assessment form for physical and environmental safety.

INTERNAL CONTROL EVALUATION FORM					
PHYSICAL AND ENVIRONMENTAL SECURITY					
No.	Name of the management area	Requirement	Control mechanism	Risk	Study Methodology
1.	INTERNAL ENVIRONMENT	Best practices	Does the University have ethical values clearly defined and appropriately adopted and communicated?	Ethical values have been inappropriately communicated	SURVEY
		Compliance with ethical values	EN ISO/IEC 27002: 2017, ITEM 11.1.5, WORK IN SAFE AREAS	No Training	
		Professional competences	Does the University conduct periodic training for employees with access to safe areas?	Verify in the Human Resources Department + interview with the HR Manager	
			EN ISO/IEC 27002: 2017, ITEM 11.1.5, WORK IN SAFE AREAS	No Training	
			Does the University conduct periodic training for employees as regards physical and environmental safety?	Verify in the Human Resources Department + interview with the HR Manager	
1.	INTERNAL ENVIRONMENT	Article 37 GDPR	Does the DPO have appropriate competences?	The DPO does not have the appropriate knowledge and experience.	VIEWS + INTERVIEW
		EN ISO/IEC 27002: 2017, ITEM 11.1.2, PHYSICAL SECURITY OF ENTRANCES	Who grants access to a given area/zone of the University? Does this person have an appropriate record in the scope of their duties?	There are no relevant provisions in the scope of duties	
		EN ISO/IEC 27002: 2017, ITEM 11.1.2, PHYSICAL SECURITY OF ENTRANCES	If the University has a register of people authorized to enter the safe zone – if so – who is responsible for keeping it? Does this person have an appropriate record in the scope of their duties?	1. Lack of records. 2. There are no relevant provisions in the scope of duties	
		Organisational structure		1. Verify the records. 2. Verify the scope of duties in the Human Resources Department	
1.	INTERNAL ENVIRONMENT	EN ISO/IEC 27002: 2017, ITEM 11.1.2, PHYSICAL SECURITY OF ENTRANCES	When the third party support staff are granted restricted access to secure areas or confidential information processing means, who grants such access? Does this person have an appropriate record in the scope of their duties?	1. No separate areas for third party entities. 2. There are no relevant provisions in the scope of duties	VERIFICATION
				Verify the scope of duties in the Human Resources Department	

INTERNAL CONTROL EVALUATION FORM							
PHYSICAL AND ENVIRONMENTAL SECURITY							
No.	Name of the management area	Name of the sub-area - more detailed	Requirement	Control mechanism	Risk	Study	Methodology
1.	INTERNAL ENVIRONMENT	Organisational structure	EN ISO/IEC 27002: 2017, ITEM 11.1.2 PHYSICAL SECURITY OF ENTRANCES	When the third party support staff are granted restricted access to secure areas or confidential information processing means, who ensures authorization?	No person monitoring the granting of restricted access. No authorizing person	Verify the scope of duties in the Human Resources Department	VERIFICATION
			EN ISO/IEC 27002: 2017, ITEM 11.2.1 LOCATION AND PROTECTION OF THE EQUIPMENT	If the University monitors environmental conditions (temperature and humidity) to detect their adverse impact on the operation of information processing measures, who monitors it? Does this person have an appropriate record in the scope of their duties?	1. No monitoring of the University in this regard. 2. There are no relevant provisions in the scope of duties	Verify the scope of duties in the Human Resources Department	
			EN ISO/IEC 27002: 2017, ITEM 11.2.2 SUPPORT SYSTEMS	Who verifies that the support systems (power supply, communication systems, ventilation and air conditioning) comply with the manufacturer's specifications?	1. No verifications carried out in this regard. 2. There are no relevant provisions in the scope of duties	Verify the scope of duties in the Human Resources Department	
			EN ISO/IEC 27002: 2017, ITEM 11.2.2 SUPPORT SYSTEMS	Who verifies that the support systems (power supply, communication systems, ventilation and air conditioning) comply with local legal requirements? How often?	1. No verification in this regard. 2. There are no relevant provisions in the scope of duties	Verify the scope of duties in the Human Resources Department	
			EN ISO/IEC 27002: 2017, ITEM 11.2.2 SUPPORT SYSTEMS	If the support systems (power, communication, ventilation and air conditioning) are regularly evaluated as for interaction with other support systems, who performs such evaluation?	There are no relevant provisions in the scope of duties	Verify the scope of duties in the Human Resources Department	

5. Conclusions

Is it worth introducing the internal audit in the organization? Indeed, an old joke says that control is the highest form of trust. An audit, however, should not be treated only as a control tool, but as a tool for managing the organization. Nowadays, the need to introduce internal audit in organizations results from the need to expand control and supervisory instruments in order to strengthen rationality, transparency and accountability for all activities undertaken in the area of management.

The result of performing the tasks of a correctly prepared and conducted audit is the improvement of the security of IT systems after the implementation of the recommendations issued by the auditor. An extremely important element is also the increase in security resulting from the implementation of the educational goal – increasing the awareness of employees and superiors in the field of safety. An extensive discussion of the encountered problems and the presentation of the arguments supporting the indicated recommendations results in a greater understanding of the necessity to introduce changes, which often results in limitations and reduced comfort of work.

The system is as safe as its weakest link is. It is the weakest elements, the least protected places, and gaps that become the most common gate to our systems, and, for this reason, an information security audit may prove so important for any organization.

References

- Czerwiński K., Książka procedur audytu wewnętrznego, CISA, Warszawa 2002, unpublished materials
- Duchniewicz S., *Metody organizacji i zarządzania*, PTM, Warszawa 2005
- Fayol H., *L'exposee des principes generaux d'administration*, 1909
- Grzyb M., Kowalik D. et al., Opracowanie metodologii zadania audytu wewnętrznego, zapewniającego w obszarze IT w kontekście bezpieczeństwa informacji, diploma paper
- International Standards for the Professional Practice of Internal Auditing, The Institute of Internal Auditors, www.ii.org.pl
- PN-ISO/IEC 27001, 2014 standard

Legal acts

The Act of 27 August 2009 on Public Finances (Polish Journal of Laws of 2019, item 2020)

Regulation of the Council of Ministers of 12 April 2012 on the National Interoperability

Framework, minimum requirements for public registers and electronic information exchange, and minimum requirements for ICT systems (Polish Journal of Laws of 2012, item 526)

Communication No. 23 of 16 December 2009 on management control standards for the public finance sector

A decorative grid pattern consisting of light gray lines and dots, arranged in a staggered, triangular-like shape, located in the upper right corner of the page.

Piotr Komsta
ORCID No. 0000-0002-0162-5518
Andrzej Frycz Modrzewski Krakow University

Prioritizing areas of implementation within the dynamic modelling concept as regards implementation processes of integrated systems

Abstract

The paper presents the importance of prioritizing the areas of implementation of IT systems supporting the management and presents the method of task implementation based on the concept of dynamic modelling. The article presents exemplary results of the analyses in this area.

Introduction

The purpose of this study is to present the importance and the manner of prioritizing the areas for the implementation of IT systems supporting management, using the dynamic modelling concept. The goal-oriented interpretation of IT system implementations is based upon improving the efficiency of the company's functioning. An accurate diagnosis of the problems (in the organizational and procedural sense) in the areas subject to implementation is the foundation for the activities that require implementation of an IT project. An accurate identification of the implementation areas of the system and prioritizing them improves the projects' chances for success. Prioritization of the areas subject to system implementation is of particular importance in the context of the goals set and the usually occurring project budget deficit. A following thesis has been adopted in the discussed study: the concept of dynamic modelling is a tool supporting prioritization as regards system implementation activities.

Studies have been carried out in this regard at one of the universities in Poland as part of pre-implementation analyses of the planned IT projects.

1. Prioritization of system implementation areas

Pre-implementation analysis is a key element of IT project planning, which allows for evaluating the system's adjustment to the company's needs¹. Planning the scope of an IT project is a difficult task and must be based on solid foundations allowing for precise definition of the problem in the context of a process. The tool that allows us to evaluate the problem in this regard is the analysis of business processes. The analysis of business processes should show the subject of the system implementation in terms of procedures, together with the efficiency assessment of the implemented procedures and include possible proposals for changes in their functioning. Only the foundations created in this way can be a basis for precise definition of the goals and the scope of the planned IT project. The efficiency assessment, which should be an element of the above analysis, allows for prioritization of the areas of system implementation which are important as regards the company's efficiency and the goals achieved.

¹ A. Bytniewski, Ł. Kristof, J. Szolc, *Zintegrowany system zarządzania Naviro*, [in:] *Zintegrowany system informatyczny. Dobre praktyki wdrożeń systemów klasy ERP*, PWN, Warszawa 2012, p. 72; R. Kowal, P. Nowak, S. Stanek, *Wdrożenie systemu SAP Business ONE w firmie produkcyjnej. Dobre praktyki wdrożeń systemów klasy ERP*, PWN, Warszawa 2012, p. 160.

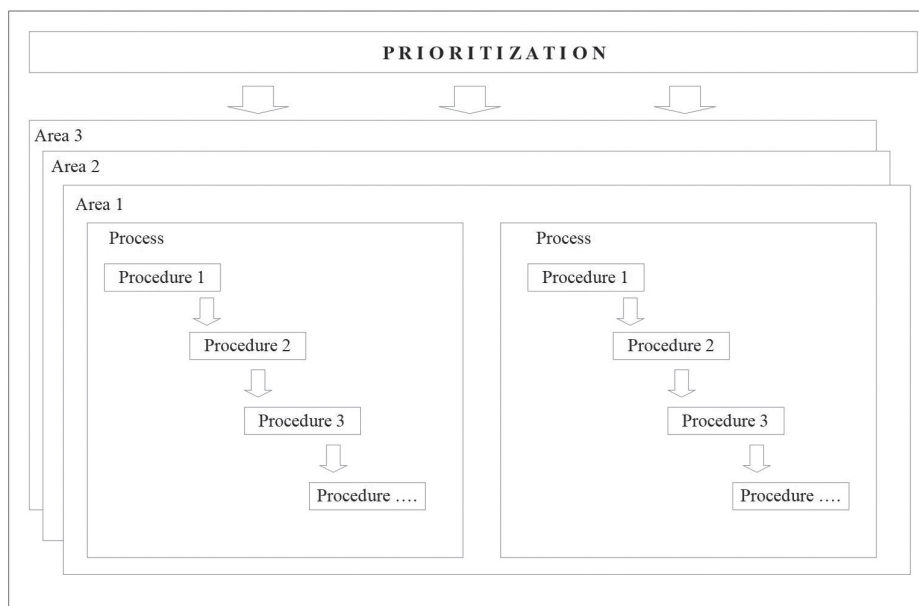


Figure 1. Prioritization of the areas of IT system implementation.

Source: own elaboration.

Defining priorities allows for rationalization of implementation activities through proper definition of the project objectives and, consequently, the planned benefits of the system implementation. The prioritization of the areas of system implementation will also be of great importance for implementation processes where, due to financial conditions, a gradation of goals should be made. Planning projects is also associated with obtaining sources of their financing. There is a problem of the proper justification of IT investments, without which obtaining financing may prove very difficult. The analysis of business processes and the prioritization of the implementation activities are a basis on which to build rational justifications for the planned IT projects.

It should be remembered that planning IT projects should go together with the realization of the company's strategic goals².

² Q. Hammouri, M. Al-Sebae, E. Abu-Shanab, Justifying the Investment of Information Technology Projects: A Case Study from Jordan, Conference Paper, April 2016, p. 58; Conference: The 15th Scientific Annual International Conference for Business "Sustainability and Competitiveness in Business" At: AL-Zaytoonah University of Jordan, https://www.researchgate.net/publication/315643289_Justifying_the_Investment_of_Information_Technology_Projects_A_Case_Study_from_Jordan (accessed on November 3, 2019).

2. Prioritization and the dynamic modelling concept

The dynamic modelling concept supports the optimization of the implementation processes for integrated systems. Basing on the analysis of the relevance and the adequacy of implementation processes, the concept allows for the development of an optimal system implementation path, taking into account the implementation requirements of an IT project. In order to learn more about dynamic modelling, the author quotes his previous studies on the above issues³. The dynamic modelling concept can be used in project-related areas while preparing a company for the planned IT projects, by making a thorough diagnosis of implementation areas, including the prioritization of the planned projects. The key element of modelling is analysing business processes and, based on the results of the analysis, identifying and prioritizing areas that are the subject to the implementation of IT systems.

The dynamic modelling concept assumes the construction of a multidimensional model of the company's operation as regards processes, taking into account pro-quality measures with respect to the planned IT projects.

As mentioned above, analytical works were carried out at one of the universities in Poland as part of pre-implementation analyses constituting an element of preparation for the planned IT projects. Selected analyses concerning the following areas are presented below: administration services for the University and student recruitment. Codification of data with regard to processes has been carried out by the University's employees basing on the questionnaires containing the data necessary for building a reference model for the University's functioning in the

³ P. Komsta, *Uwarunkowania i obszary modelowania dynamicznego procesów implementacyjnych systemów zintegrowanych*, [in:] *IT w organizacjach gospodarczych. Wybrane zagadnienia*, Dom Organizatora, Toruń 2010, pp. 39–45; P. Komsta, *Kształtowanie procedur implementacyjnych systemów zintegrowanych w koncepcji modelowania dynamicznego*, [in:] *Narzędzia informatyczne w gospodarce elektronicznej i systemach wspomagania decyzji*, Wydawnictwo Politechniki Częstochowskiej, Częstochowa 2011, pp. 88–93; P. Komsta, *Mapowanie i analiza parametrów w dynamicznym modelu implementacji systemów zintegrowanych*, [in:] *Wykorzystanie wybranych technologii komunikacji w zarządzaniu wartością organizacji*, Wydawnictwo Politechniki Częstochowskiej, Częstochowa 2012, pp. 277–282; P. Komsta, *Punkty węzłowe w modelowaniu dynamicznym procesów implementacyjnych systemów zintegrowanych*, [in:] *Technologie informacyjne w funkcjonowaniu organizacji. Zarządzanie z wykorzystaniem multimediów*, Dom Organizatora, Toruń 2013, pp. 521–528; P. Komsta, *Czynniki sprawności modelowania dynamicznego procesów implementacyjnych systemów zintegrowanych*, [in:] *Wybrane zastosowania technologii informacyjnych wspomagających zarządzanie w organizacjach*, Wydawnictwo Politechniki Częstochowskiej, Częstochowa 2015, pp. 42–50; P. Komsta, *Cele modelowania dynamicznego procesów implementacyjnych systemów zintegrowanych*, [in:] *Innowacje i przedsiębiorczość ujęcie makro- i mikroekonomiczne*, Wydawnictwo Naukowe Wyższej Szkoły Biznesu w Dąbrowie Górniczej, Dąbrowa Górnicza 2016, pp. 189–196.

above-mentioned areas. The questionnaires were also supposed to estimate the functioning of the procedures in terms of efficiency⁴ as well as possible proposals for organizational changes. As a result of data codification and parameter mapping, a multidimensional reference model of the University's functioning was created, on whose basis the functional requirements for the IT systems, being the subject of the planned IT projects was built. The construction of functional requirements of the systems was carried out with the use of a matrix of connections, which allowed for combining the planned system functionalities with the procedures of the constructed reference model.

The constructed model allowed for multidimensional analyses which helped in identifying the procedures governing the implementation of individual processes subject to system implementation in individual domain areas. It also allowed for estimating the amount of work required in the above areas, as well as for indicating the procedures to which special attention should be paid when undertaking design activities in the context of planned IT projects.

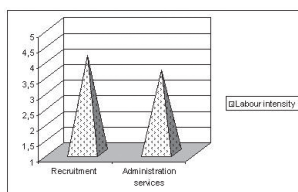


Figure 2. The amount of work required in the area of University administration services and recruitment areas.

Source: own elaboration.

Figure 2 shows the average values determining the amount of work necessary in the areas in question. The above analysis indicated that the tasks carried out refers to both office services for the University and the area of student recruitment which involved large forces and resources. Thus, the results of the analyses corroborated the reasonableness of implementing IT investments in the above areas. The analysis also indicated that the amount of work involved in recruiting students was much higher than in the area of University administration services, constituting the basis for possible prioritization activities in this regard.

⁴ The scale of the level of labour consumption was determined in the range from 1 to 5, where 1 meant a procedure with very low labour intensity and 5 a procedure with very high labour intensity.

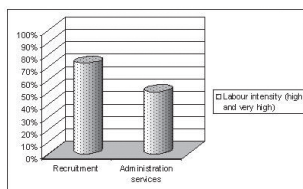


Figure 3. Comparative analysis of the percentage of highly and very highly labour-intensive procedures in the areas of University administration services and recruitment process.

Source: own elaboration.

More detailed analyses indicated large differences between these areas in terms of the share of highly and very highly labour-intensive works, as well as the heterogeneity of labour intensity in both areas.

Figure 3 presents a comparative analysis of both areas in terms of the share of highly and very highly labour-intensive works.

The share of highly and very highly labour-intensive work was much higher in the recruitment area than in the area of University's administration services.

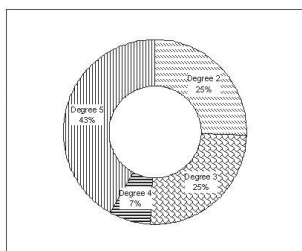


Figure 4. Percentage of the procedures due to the level of labour intensity in the area of University administration services.

Source: own elaboration.

Figures 4 and 5 show examples of analyses concerning the degree of labour-intensity for the procedures in both of the areas separately. The analyses confirmed the heterogeneity of the labour intensity of the procedures in each of the exemplary areas.

In case of the University administration work, the number of procedures with a very high and high degree of labour intensity was identical to the number of procedures with a medium and low degree of labour intensity. At the same time, it should be emphasized that the share of the procedures with a medium degree of labour intensity in the area of administration services was much higher than for the works related to student recruitment.

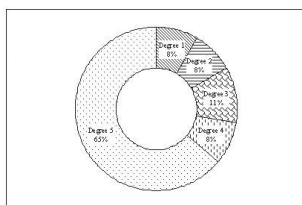


Figure 5. The percentage of the procedures according to the level of labour intensity for the University administration services.

Source: own elaboration.

In the case of handling the recruitment process, very high and highly labour-intensive procedures prevailed. Detailed analyses also allowed for setting priorities as to possible project-based activities related to individual procedures.

The construction of the reference model allowed not only for identifying the problems, but also formed a basis for the proposed solutions. It should be emphasized that the foundation for created database were the proposals of changes and improvements submitted by employees responsible for the implementation of individual procedures. This knowledge was helpful in building the functional requirements specification of the system. Building the specification of the system's functional requirements requires a thorough analysis of problems in the context of rationalizing the implementation activities with regard to the system, and thus improving the efficiency of the company's functioning. The use of the dynamic modelling concept, involving the construction of a reference model of the company's functioning based on the performed analyses, allowed for prioritization activities and thus increased awareness with respect to the rational selection of the organizational, functional and technological solutions aimed at improving the efficiency of the University functioning.

Conclusions

The process of investment in IT technologies should correspond to the company's business and should have an impact on the results achieved by the company⁵. Planning of IT investments must be based on solid foundations allowing for rationalizing the activities in this regard. The planned IT investments must consider the business aspect of the company's goals and respond to related challenges. Meeting this requirement involves a thorough diagnosis of the problems and setting priorities in investment. In order to meet market requirements, the company must develop dynamically in the constantly changing business environment, and thus must dynamically adapt its goals and tools to the implementation of these goals. We might say that the process of investment in information technologies is a continuous process and should correspond to the market challenges that the company has to face on an ongoing basis. The use of the dynamic modelling concept in the planning of IT projects made it possible to correctly identify the areas for system implementation in terms of processes, as well as to diagnose the problem in terms of efficiency, to combine the process aspect of the University's functioning with the technological aspect and to define priorities in the scope of system implementation activities.

References

- Bytniewski A., Kristof Ł., Szolc J., *Zintegrowany system zarządzania Naviro*, [in:] *Zintegrowany system informatyczny. Dobre praktyki wdrożeń systemów klasy ERP*, PWN, Warszawa 2012
- Hammouri Q., Al-Sebae M., Abu-Shanab E., *Justifying the Investment of Information Technology Projects: A Case Study from Jordan*, Conference Paper, April 2016; Conference: The 15th Scientific Annual International Conference for Business "Sustainability and Competitiveness in Business" At: AL-Zaytoonah University of Jordan, https://www.researchgate.net/publication/315643289_Justifying_the_Investment_of_Information_Technology_Projects_A_Case_Study_from_Jordan (accessed on November 3, 2019)
- Komsta P., *Uwarunkowania i obszary modelowania dynamicznego procesów implementacyjnych systemów zintegrowanych*, [in:] *IT w organizacjach gospodarczych. Wybrane zagadnienia*, Dom Organizatora, Toruń 2010, pp. 39–45

⁵ L. Haewon, C. Hanbyeol, L. Junyeong, M. Jinyoung, L. Heeseok, *Impact of IT Investment on Firm Performance Based on Technology IT Architecture*, "Procedia Computer Science", 91 (2016), p. 660.

- Komsta P., *Kształtowanie procedur implementacyjnych systemów zintegrowanych w koncepcji modelowania dynamicznego*, [in:] *Narzędzia informatyczne w gospodarce elektro-nicznej i systemach wspomagania decyzji*, Wydawnictwo Politechniki Częstochowskiej, Częstochowa 2011, pp. 88–93
- Komsta P., *Mapowanie i analiza parametrów w dynamicznym modelu implementacji systemów zintegrowanych*, [in:] *Wykorzystanie wybranych technologii komunikacji w zarządzaniu wartością organizacji*, Wydawnictwo Politechniki Częstochowskiej, Częstochowa 2012, pp. 277–282
- Komsta P., *Punkty węzłowe w modelowaniu dynamicznym procesów implementacyjnych systemów zintegrowanych*, [in:] *Technologie informacyjne w funkcjonowaniu organizacji, Zarządzanie z wykorzystaniem multimediiów*, Dom Organizatora, Toruń 2013, pp. 521–528
- Komsta P., *Czynniki sprawności modelowania dynamicznego procesów implementacyjnych systemów zintegrowanych*, [in:] *Wybrane zastosowania technologii informacyjnych wspomagających zarządzanie w organizacjach*, Wydawnictwo Politechniki Częstochowskiej, Częstochowa 2015, pp. 42–50
- Komsta P., *Cele modelowania dynamicznego procesów implementacyjnych systemów zintegrowanych*, [in:] *Innowacje i przedsiębiorczość ujęcie makro- i mikroekonomiczne*, Wydawnictwo Naukowe Wyższej Szkoły Biznesu w Dąbrowie Górniczej, Dąbrowa Górnicza 2016, pp. 189–196
- Kowal R., Nowak P., Stanek S., *Wdrożenie systemu SAP Business ONE w firmie produkcyjnej. Dobre praktyki wdrożeń systemów klasy ERP*, PWN Warszawa 2012
- Haewon L., Hanbyeol C., Junyeong L., Jinyoung M., Heeseok L., *Impact of IT Investment on Firm Performance Based on Technology IT Architecture*, “Procedia Computer Science”, 91 (2016), pp. 652–661



Index of names

Abu-Shanab Emad 107

Al-Sebae Mai 107

Babik Wiesław 72

Balcerak Alicja 56, 69

Beaverstock Malcolm 55

Beynon-Davies Paul 18, 53

Błaszczyk Katarzyna 54

Butcher Tim 52, 68

Buregwa-Czuma Sylwia 11

Bytniewski Andrzej 106

Cabała Paweł 73

Chuchro Ewa 73

Cukier Kenneth 38

Czerwiński Krzysztof 85

Drucker Peter F. 84

Duchniewicz Stanisław 84

Einstein Albert 19

Fayol Henri 84

Gałęzowski Dariusz 53, 54

Garwol Katarzyna 11

Gleick James 14

Goban-Klas Tomasz 11, 15

Greenwood Allen 55

Haewon Lee 112

Hammouri Qais 107

Hanbyeol Choi 112

Heeseok Lee 112

Janicki Wojciech 54, 55, 63, 68

Januszko-Szakiel Aneta 72–74, 79

Jinyoung Min 112

Jones William P. 73

Junyeong Lee 112

Kisielnicki Jerzy 11

Knosala Ryszard 54

Komsta Piotr 108

Korycińska Paloma 72, 74, 79

Kowal Radosław 106

Kristof Łukasz 106

Kubiatowicz John 19

Kuraś Marian 12

Kwaśnicki Witold 56, 69

Lalwani Chandra 52, 68

Lansdale Mark W. 73

Lavery Eamonn 55

Le Bellac Michel 19
Lem Stanisław 18, 19

Mangan John 52, 68
Marshall Catherine C. 74
Martyniak Zbigniew 51, 68
Masuda Yoneji 10, 11, 14–16
Materska Katarzyna 72, 73
Mayer-Schonberger Viktor 38
Moryś Adriana 11

Nordgren William 55
Nowak Jerzy S. 11
Nowak Piotr 106

Pietruch-Reizes Diana 73

Redwine Gabriela 74, 79

Sapa Remigiusz 73
Schmitt Eric 19
Semenov Iouri N. 52
Sienkiewicz Mariusz 53, 54
Skrzypek Elżbieta 68
Sosińska-Kalata Barbara 73
Stanek Stanisław 106
Steinbeck Hans H. 68
Świgoń Marzena 73
Szolc Jarosław 106

Teevan Jaime 73
Tomaszczyk Jacek 73
Tyrańska Małgorzata 73

Wiktorowska-Jasik Anna 52, 68
Wilkowski Marcin 74

The following issues can be clearly noted in this monograph: information-based civilization, i.e. defining civilization through information, the discussion of the concept of information, language as an information carrier, the role of information in private life and in business, information security, archiving and management.

In all considerations, the Authors pay attention to the important role of information in making management decisions. The issues related to the core, processing, security and archiving of information are considered from the perspective of civilization, organization (enterprise) and from a personal point of view.



Fundusze Europejskie
Wiedza Edukacja Rozwój



**Rzeczpospolita
Polska**

Unia Europejska
Europejski Fundusz Społeczny



ISBN 978-83-65548-82-5